



Universidade de Aveiro Departamento de Biologia
Ano 2009

**Cristina Sofia de
Sousa Cardoso e
Valente dos Santos**

**Os Ciganos de Portugal
Uma perspectiva genética da sua história**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biologia Molecular e Celular, realizada sob a orientação científica da Doutora Maria João Prata Martins Ribeiro, Professora associada com agregação da Faculdade de Ciências da Universidade do Porto.

Apoio financeiro da Fundação da
Ciência e Tecnologia através do
projecto PTDC/ANT/70413/2006 e do
Programa Operacional para a Ciência e
Tecnologia e Inovação (POCI - 2010).

o júri

presidente

Prof. Doutora Maria de Lourdes Gomes Pereira

Professora Associada com Agregação do Departamento de Biologia da Universidade de Aveiro

Vogais:

Prof. Doutora Maria João Prata Martins Ribeiro (orientadora)

Professora Associada com Agregação da Faculdade de Ciências da Universidade do Porto

Prof. Doutora Maria Adelaide de Pinho Almeida (co-orientadora)

Professora Auxiliar do Departamento de Biologia da Universidade de Aveiro

Prof. Doutora Cristina Maria Pereira Santos (arguente)

Professor Ajudante Doutor do Departamento de Biologia Animal, Biologia Vegetal e Ecologia da Faculdade de Biociências da Universidade Autónoma de Barcelona

agradecimentos

Em primeiro lugar, e com todo o respeito às pessoas que irei citar por tornarem possível a realização e a escrita deste trabalho, gostaria de mostrar a minha gratidão para com a Professora Maria João, que não só me confiou a continuação deste projecto na área da Genética Populacional, como também se mostrou sempre disponível, prontificando desde início a sua ajuda, de forma sincera e verdadeira.

Ao Professor António Amorim pela oportunidade em poder fazer parte integrante do seu grupo de investigação.

À Leonor, o meu obrigado, que sem qualquer obrigatoriedade académica, foi sem dúvida a pessoa que mais me fez “crescer” em termos científicos e profissionais durante a minha estadia no IPATIMUP.

À disponibilidade contínua da Prof. Adelaide, que me vem acompanhando desde a minha licenciatura, e me encorajou a prosseguir esta “longa” caminhada.

Aos meus novos colegas, que me receberam de foram ímpar (Iva, Ana Goios, Rui, Inês, Sofia e Rita), um obrigado, por se mostrarem sempre disponíveis em me ajudar, e me concederem momentos divertidos durante os nossos almoços.

Às minhas colegas de gabinete, Nádia e Cíntia, obrigada pelas risadas e descontração proporcionadas.

Aos restantes elementos do GEPO, por bem me terem acolhido no grupo de trabalho.

Aos meus amigos do peito (vocês sabe quem são) que estiveram, estão e estarão sempre a meu lado.

Aos meus pais, Dédinho e mãe, um beijo pelo apoio.

Mas é sem dúvida à minha mana a quem dedico este trabalho, porque é muito mais do que um laço de sangue que nos une...

palavras-chave

ciganos portugueses, Índia, Europa, Genética Populacional, microssatélites, autossomas, cromossoma X.

resumo

No presente trabalho caracterizou-se a população cigana portuguesa quanto a microssatélites (STRs) localizados nos autossomas e no cromossoma X, com o principal objectivo de inferir, através da análise genética, alguns aspectos da sua história. Para o efeito estudou-se uma amostra de 127 indivíduos não aparentados, pertencentes a 18 comunidades ciganas, distribuídas por 11 distritos do país. Tanto em STRs autossómicos como do cromossoma X, os ciganos apresentaram níveis mais baixos de diversidade intrapopulacional que a população portuguesa não-cigana. A redução de diversidade pode ser explicada pela acção combinada dos efeitos de deriva induzidos pelo pequeno tamanho populacional, e pelas práticas endogâmicas tradicionais executadas pela maioria das populações ciganas. A avaliação da diferenciação interpopulacional revelou que as distâncias médias entre populações ciganas eram muito superiores às registadas entre populações europeias não-ciganas. Este resultado traduz novamente os fortes efeitos de deriva que modelaram o perfil genético das populações ciganas, indiciando ainda uma considerável limitação do fluxo génico entre populações ciganas e respectivas populações hospedeiras, o que também se compreende dadas as suas tradições sócio-culturais.

No contexto da diversidade genética europeia, as populações ciganas tendem a agrupar-se entre si, e a coerência do grupo evidencia claramente a partilha de uma origem recente comum.

Quando com base nos STRs autossómicos foram inferidas proporções de mistura em diferentes grupos Roma, em todos se verificou estar consideravelmente diluído o componente atribuível à origem ancestral na Índia. Comparativamente a este, o componente europeu teve tendência a ser superior, atingindo o valor mais elevado entre os ciganos portugueses. Portanto, não obstante algumas barreiras ao fluxo génico, foram substanciais nos grupos Roma os aportes genéticos resultantes de mistura com populações europeias.

Foi analisado ainda o padrão de desequilíbrio de ligação (LD) no cromossoma X. Na população cigana portuguesa detectaram-se algumas associações significativas entre STRs e um nível de LD mais elevado do que o encontrado na população portuguesa não-cigana. O padrão de LD nos ciganos enquadra-se bem na sua história marcada por mistura populacional e efeitos de deriva.

Em suma, a análise de marcadores moleculares permitiu inferir alguns aspectos importantes e reforçar outros da história demográfica dos ciganos portugueses.

keywords

Portuguese Gypsies, India, Europe, Population Genetics, autosomes, X chromosome and microsatellites.

abstract

In this study the Portuguese Gypsies were characterized for autosomal and X-chromosome microsatellites, with the aim of inferring some aspects of their population history. Viewing that, a sample of 127 unrelated individuals belonging to 18 different communities spread over 11 districts from Portugal was studied.

For both for autosomal and X-chromosome STRs, Gypsies showed reduced levels of intra-population diversity compared to the non-Gypsy Portuguese population. Reduced diversity can be explained by the combined effects of genetic drift, induced by small population size, and endogamic practices which are traditional in the majority of Roma groups.

The evaluation of inter-population differentiation revealed that the average distances between Gypsy populations were much higher than the observed between non-Gypsy European populations. Such result reflects again that strong drift effects have modulated the genetic profile of Gypsy groups, and further evidences some limitation to gene flow between Gypsies and their host populations, being also understandable given their socio-cultural traditions.

In the context of the European diversity, all Gypsy groups tend to cluster together. The coherence of the cluster clearly represents a signature of the common origin of the Gypsies.

When autosomal STRs were used to estimate admixture proportions in different Roma groups, the component ascribed to the ancestral origin in India was considerably diluted in every group. Comparatively, the European ancestrality was slightly higher, demonstrating that despite some barriers to gene flow, genetic inputs resulting from admixture with Europeans contributed substantially to the gene pool of current Roma.

It was also addressed the pattern of linkage disequilibrium (LD) in the X chromosome. Within the Portuguese Gypsies some significant associations between STRs were detected and the level of LD was higher than the observed within the Portuguese non-Gypsies. This pattern of LD fits quite well in the Gypsy history of population admixture and strong drift effects.

In summary, the analysis of molecular markers has provided important insights into the demographic history of the Gypsies.

Índice

Financiamento	ii
Agradecimentos	iv
Resumo	v
Abstract.....	vi
Índice de figuras	ix
Índice de tabelas	xii
Índice de siglas e abreviaturas	xiv
1. INTRODUÇÃO.....	16
1.1. Os ciganos europeus	16
1.1.1. Origens, história e diáspora	18
1.1.2. Os ciganos portugueses	23
1.2. Estudos em genética populacional e marcadores genéticos	25
1.2.1. Microsatélites ou <i>Short Tandem Repeats</i>	27
1.3. Cromossoma X e autossomas	29
1.3.1. STRs do cromossoma X e autossómicos	30
1.4. Objectivos.....	33
2. MATERIAL E MÉTODOS.....	34
2.1. Amostragem e material biológico.....	34
2.2. Extracção de DNA.....	35

2.3. Análise de microssatélites autossômicos e do cromossoma X.....	36
2.3.1. Amplificação por PCR multiplex.....	36
2.3.2. Genotipagem.....	41
2.4. Análise de dados.....	44
2.4.1. Análise de marcadores STRs autossômicos	44
2.4.2. Análise de marcadores X-STRs.....	45
 3. RESULTADOS E DISCUSSÃO	 48
3.1. Análise de STRs autossômicos.....	48
3.1.1. Diversidade genética.....	48
3.1.2. Comparação com diferentes populações ciganas e não-ciganas	56
3.1.3. Estimativas de miscigenação nas populações ciganas.....	66
3.2. Análise de X-STRs	70
3.2.1. Diversidade genética.....	70
3.2.2. Padrão de desequilíbrio de ligação (LD)	76
3.2.3. Diversidade interpopulacional.....	84
 4. CONSIDERAÇÕES FINAIS E CONCLUSÕES.....	 86
 5. REFERÊNCIAS BIBLIOGRÁFICAS	 88
 Anexo 1. Artigo aceite na revista <i>American Journal of Physical Antropology</i> 2009	 100
Anexo 2. 6th ISABS Conference on Human Genome Project Based Applications in Forensic Science, Anthropology and Individualized Medicine, Split, Croácia, 1-5 Junho 2009.....	101
Anexo 3. 23 rd World Congress of International Society for Forensic Genetics, Buenos Aires, Argentina, 14-18 Setembro 2009	102

Índice de figuras

Figura 1.	A dispersão das populações ciganas a partir da Índia em direcção à Europa Ocidental (<i>imagem in www.vurdon.it/itinerario.jpg</i>)	21
Figura 2.	Localização dos 10 STRs do cromossoma X estudados neste trabalho (a negrito) e respectivos grupos de ligação com indicação dos STRs flanqueantes, de acordo com Szibor <i>et al.</i> (2003a).....	32
Figura 3.	Mapa de Portugal indicando a localização das zonas de colheita das amostras dos 127 indivíduos da população cigana portuguesa	34
Figura 4.	Diagrama esquemático do multiplex usado para amplificar 15 STRs autossómicos, onde se apresentam as cores atribuídas a cada microsatélite e a gama de variação do tamanho esperado, em pares de bases (bp), para os alelos de cada STR.....	36
Figura 5.	Representação esquemática do multiplex usado para genotipar os 10 X-STRs, com referência aos intervalos de tamanho esperados para os alelos de cada STR, em pares de bases (bp), e a marcação fluorescente usada em cada microsatélite (6-FAM, azul; VIC, verde e, NED, amarelo).....	39
Figura 6.	Electroforegrama do perfil de um indivíduo obtido com o <i>kit</i> AmpF/STR Identifiler.....	42
Figura 7.	Electroforegrama do perfil de um indivíduo do sexo masculino obtido com o sistema X-decaplex	43

Figura 8.	Padrão de decréscimo de diversidade genética entre as populações ciganas e correspondentes populações hospedeiras, com a indicação da sua localização geográfica, representado para o <i>set 1</i> (4 STRs)	54
Figura 9.	Mapa da Índia com a localização geográfica (marcadas com círculo preto) das populações usadas para comparação com a população cigana portuguesa em estudo..	59
Figura 10.	<i>Plot</i> resultante da Análise Multidimensional (MDS) baseada em F_{ST} entre diversas populações Roma e não-Roma.....	62
Figura 11.	Árvore obtida por <i>Neighbour Joining</i> a partir da matriz de distâncias genéticas da tabela 12. O tamanho dos ramos é proporcional à distância genética entre populações, e a escala da figura é dada em unidades de F_{ST}	64
Figura 12.	<i>Plot</i> resultante da Análise de Componentes Principais (PCA) baseada em frequências alélicas de 13 STRs autossómicos, incluindo diversas populações Roma e não-Roma.....	65
Figura 13.	Padrões de LD na população portuguesa cigana (A) e não-cigana (B) baseados nos valores de D' . Na escala dos eixos encontram-se os 10 X-STRs analisados, tendo em conta as respectivas distâncias físicas em Mb (tabela 6). O <i>software</i> Gold representa graficamente cada par de <i>loci</i> por cores: vermelho, verde e azul significando grau forte, moderado e fraco, respectivamente de LD	83

Índice de tabelas

Tabela 1.	Características dos marcadores STR autossômicos estudados neste trabalho.....	37
Tabela 2.	Perfil de STRs autossômicos da amostra de DNA (9947A) usada como controlo positivo descrito no manual do <i>kit</i> de PCR AmpF/STR Identifiler (<i>AB Applied Biosystems</i>).....	38
Tabela 3.	Sequências dos <i>primers</i> usados e respectiva marcação com fluorocromos	39
Tabela 4.	Características dos marcadores X-STR incluídos no sistema multiplex utilizado neste trabalho	40
Tabela 5.	Perfil de X-STRs da amostra 9947A usada como controlo positivo	41
Tabela 6.	Marcadores STR descritos e sua localização física e genética no cromossoma X .	46
Tabela 7.	Frequências alélicas, heterozigotia observada (H_{OBS}) e heterozigotia esperada (H_{ESP}), valores de p obtidos no teste de Equilíbrio de Hardy-Weinberg para os 15 STRs autossômicos na população cigana portuguesa ($N=123$)	50
Tabela 8.	Valores de diversidade genética <i>locus a locus</i> para a população cigana portuguesa e outras populações ciganas europeias, e respectivas populações não-ciganas.....	51

Tabela 9.

Valores médios da diversidade genética para cada população cigana e respectiva população não-cigana, considerando três conjuntos distintos de STRs autossômicos *set 1*: restringe-se a 4 STRs (D8S1179, D3S1358, D13S317 e D5S818); *set 2*: engloba os 13 marcadores considerados para o sistema CODIS e *set 3*: inclui os 15 marcadores disponíveis no *kit* AmpF/STR Identifier 52

Tabela 10.

Valores de distância genética (F_{ST}) *locus a locus* entre a população cigana portuguesa e as cinco populações ciganas europeias, e respectivos valores de significância estatística (p), para um desvio padrão $\leq 0,001$ 57

Tabela 11.

Valores de distância genética (F_{ST}) *locus a locus* entre a população cigana portuguesa e as cinco populações não-ciganas europeias, e respectivos valores de significância estatística (p), para um desvio padrão $\leq 0,001$ 58

Tabela 12.

Valores de F_{ST} (abaixo da diagonal) entre as populações comparadas para os 13 STRs (excepto D2S1338 e D19S433), e respectivos valores de significância estatística (p) (acima da diagonal), para um desvio padrão $\leq 0,001$ 60

Tabela 13.

Estimativas de miscigenação e respectivo valor padrão (σ) para as populações ciganas europeias considerando três idades diferentes (700, 1000 e 1700 anos) ... 68

Tabela 14.

Frequências alélicas e valores de diversidade génica (H_{ESP}) em 10 X-STRs na população cigana portuguesa ($N=123$)..... 71

Tabela 15.

Valores de diversidade genética *locus a locus* e valores médios para a população cigana portuguesa e outras populações com dados publicados..... 73

Tabela 16.

Valores de p de testes exactos de desequilíbrio de ligação entre pares de *loci*, respectivo desvio padrão (σ) e parâmetro D' , para os alelos mais frequentes de cada par de marcadores, nas populações de portugueses ciganos e não-ciganos.

A distância genética entre marcadores apresenta-se em centiMorgans (cM) 77

Tabela 17.

Frequências haplotípicas e respectivos valores de desvio padrão (σ) envolvendo os microsatélites DXS9898, DXS6809 e DXS6789, na população cigana portuguesa 80

Tabela 18.

Valores de F_{ST} entre a população cigana portuguesa e outras populações para os 10 X-STRs e respectivos valores de significância estatística (p), para um desvio padrão $\leq 0,001$ 85

Lista de siglas e abreviaturas

bp - pares de bases

CGM - *Continuous Gene Flow*

cM - centiMorgan

CODIS - *Combined DNA Index System*

DNA - ácido desoxirribonucleico

HI - *Hybrid Isolation*

H_{ESP} - heterozigotia esperada

H_{OBS} - heterozigotia observada

IAMM - *Infinite Allele Mutation Model*

ISFG - *International Society of Forensic Genetics*

LD - desequilíbrio de ligação

Mb - mega pares de bases

MDS - *Multidimensional Scalling*

mg - miligrama

min - minutos

mL, µL - mililitro, microlitro

mtDNA - DNA mitocondrial

PCR - *Polymerase Chain Reaction*

rpm - rotações por minuto

RFLP - *Restriction Fragment Length Polymorphism*

seg - segundos

SMM - *Stepwise Mutation Model*

SNP - *Single Nucleotide Polymorphism*

STR - *Short Tandem Repeat*

Xp - tel - braço curto do cromossoma X

X-STR - STR localizado no cromossoma X

1. Introdução

1.1. Os ciganos europeus

O termo “cigano” não tem origem etimológica no *romani*, a língua falada pelos ciganos, mas sim na palavra grega medieval *atsinganos* ou *atsinkanos*, que significa “intocáveis”, e que era o nome dado na altura a um grupo bizantino cujos elementos eram conhecidos pelas suas artes mágicas e de encantadores de serpentes, e que se ocupavam em ler a sina (Bloch 1960, *in* Costa, 2003). A mesma origem parece ter designações usadas noutros países para referir os ciganos, como *tsigane* em França, *zigeuner* na Alemanha ou *zingari* em Itália. Já os termos *gitano* aplicado em Espanha aos *kalo* os grupos de ciganos da Península Ibérica, ou *gypsie* usado em Inglaterra, podem ser reminiscências de uma lenda antiga segundo a qual os primeiros ciganos que chegaram à Europa seriam originários do Egipto (Costa, 2003).

Actualmente, o termo *Roma* é cada vez mais utilizado, principalmente quando se faz referência aos ciganos da Europa de Leste (Kalaydjieva *et al.*, 2005).

Para Liégeois (1989), é pouco adequado usar o termo “cigano” no singular para referir os ciganos em geral, pois mascara um dos seus aspectos estruturais mais marcantes: a diversidade de grupos. Por essa razão, e de forma a integrar esse elemento, será mais apropriado falar antes em populações “ditas ciganas” que, ainda nas palavras de Liégeois, constituem “um mosaico fluido de grupos diversificados” (Costa, 2003).

Na realidade, os ciganos europeus englobam vários subgrupos, entre os quais os *Lovara* e *Kalderas*, estabelecidos na Europa Central e Oriental (do grupo *Rom*, dito *Vlax* ou *Valaque*), os *Sinti* e os *Manouche* da Europa Ocidental (do grupo *Sinto-Mãnus*), os *Valstiké* (um dos sub-grupos *Manouches*), ou os *Kalé*, em que se incluem os ciganos de Portugal, Espanha e do Sul da França, não fazendo referência a grupos mais pequenos (Fraser, 1998).

Em certa medida os ciganos contrariam a noção convencional de população europeia: não pertencem propriamente a uma nação/estado, falam línguas variadas e as práticas religiosas diferem de grupo para grupo (Kalaydjieva *et al.*, 2001a).

Na Europa, os ciganos estão representados por cerca de 8 a 10 milhões de indivíduos. Embora espalhados por todo o continente, as maiores concentrações encontram-se em zonas do sudeste europeu e Península Ibérica (Kalaydjieva *et al.*, 2001a).

Alguns vestígios históricos indicam que os primeiros ciganos podem ter chegado à Europa tão cedo quanto há cerca de 1000 anos atrás, entrando pela região sul dos Balcãs (Fraser, 1998).

O aparecimento de numerosos grupos tão diferentes entre si resultou de um processo de dispersão pela Europa, que durou séculos e envolveu sucessivas migrações sistematicamente acompanhadas de episódios de fragmentação populacional (Reyniers 1995, *in* Costa, 2003). Pequeno tamanho populacional, propiciador de fortes efeitos de deriva genética, fluxo genético limitado entre grupos e mistura diferencial entre ciganos e não-ciganos, foram alguns dos factores que conduziram a uma rápida diferenciação entre grupos Roma, ou seja a um padrão de substruturação muito marcado, sem paralelo com o que se encontra geralmente entre outras populações europeias (Gresham *et al.*, 2001; Kalaydjieva e Morar, 2003). Daí que em média sejam maiores as distâncias genéticas entre diferentes grupos Roma do que entre populações europeias não-ciganas (Kalaydjieva *et al.*, 2001a, Gusmão *et al.*, 2008). Por outro lado, ao contrário do que se observa nas populações europeias em que é elevada a correlação entre distância geográfica e diferenciação genética, entre os Roma o factor geográfico pouca influência parece ter no grau de diferenciação intergrupos (Gresham *et al.*, 2001; Gusmão *et al.*, 2008).

A endogamia, uma prática comum nas populações ciganas, é outra causa que contribui para aumentar a diferenciação entre grupos ao mesmo tempo que reduz os níveis de diversidade intrapopulacional. Comparativamente às respectivas populações hospedeiras, todos os grupos Roma se caracterizam por níveis baixos de diversidade genética (Kalaydjieva *et al.*, 2005). As tradições sócio-culturais têm representado uma certa limitação ao fluxo genético entre grupos Roma e populações vizinhas, quer sejam ciganas ou não. Os Roma tendem a manter alguma rigidez nas regras matrimoniais, evitando casamentos com indivíduos de outras populações, inclusivé de outros grupos

ciganos, dando preferência aos que ocorrem dentro do próprio grupo (Liégeois, 1989; Kalaydjieva *et al.*, 2005).

Mas como alguns autores salientam (Marushiakova e Popov, 1997; Fraser, 1998) o “grupo” funciona como a unidade primária da organização social nas populações ciganas e, sendo assim, nenhum grupo cigano pode ser representativo de outros grupos Roma (Fraser, 1998).

1.1.1. Origens, história e diáspora

A cultura *rom*, cigana, e a sua língua, o *romani*, constituem antes de mais uma curiosa realidade etnográfica, pois embora partilhem a mesma matriz indo-europeia que os seus hóspedes europeus, os ciganos apresentam uma organização social muito distinta, irreduzível à morfologia social predominante na Europa, e em que muitos vêm sinais de uma ancestralidade na Índia (Perez, *in* Coelho, 1995).

Também falam uma língua diferente de qualquer outra falada no continente europeu, tendo sido as evidências linguísticas, que logo no século XVIII, primeiro apontaram para a Índia como local de origem das populações de que derivaram os ciganos actuais. Estudos linguísticos posteriores corroboraram essa hipótese, mas contudo, a ausência de afinidades particulares entre o *romani* e algumas das línguas ou dialectos actualmente falados no sub-continento indiano, não têm auxiliado a esclarecer qual o grupo populacional de que saíram os *proto-Roma*, ou se estes tiveram origem num único ou em diversos grupos ancestrais indianos (Turner 1926, Sampson 1927 e Hancock 2000, *in* Costa, 2003).

Posteriormente, a origem indiana tem sido fortemente suportada por estudos genéticos de diversa índole. De entre os mais recentes, destacam-se os que têm recorrido a sistemas não-recombinantes (DNA mitocondrial ou cromossoma Y), e que têm vindo a demonstrar que todos os grupos Roma compartilham alguns tipos de linhagens, masculinas e femininas, que habitualmente não se encontram em populações da Europa ou Médio Oriente, mas pelo contrário, são comuns em populações indianas. Parecem representar

linhagens fundadoras de origem indiana, cuja presença em todos os grupos Roma evidencia a sua origem comum (Gresham *et al.*, 2001; Gusmão *et al.*, 2008).

Se a origem na Índia é aceite com razoável consenso, sabe-se muito pouco sobre os contornos da diáspora já que as fontes históricas são muito vagas e frequentemente contraditórias.

A saída dos primeiros ciganos da Índia, considerada como a primeira grande dispersão dos ciganos, pode ter ocorrido ainda em meados do século III, na sequência da conquista do norte da Índia pelos Persas (Liégeois, 1989; Kenderick e Puxon, 1998). A pressão da invasão persa, terá impulsionado a primeira vaga de migração da Índia em direcção à Pérsia, numa busca por melhores condições de trabalho e menos convulsão social (Costa, 2003). Contudo, há quem defenda que os ciganos provieram de um grupo de *Rajputs*, que se denominavam a si próprios de *Romane Chav* (filhos de Rama) e que, escapando às invasões dos turcos e mongóis, ingressaram do Afeganistão, deslocando-se depois para a Grécia (Kochanowski, *in* Costa, 2003).

Nada indica que a emigração inicial da Índia tenha constituído uma movimentação massiva de indivíduos. Se tal se tivesse verificado, não se compreenderia a total ausência de relatos históricos sobre o sucedido, pelo que se tem admitido que o número de indivíduos envolvidos na migração inicial tenha sido relativamente reduzido. Assim, em termos de efeitos populacionais, a saída da Índia acompanhou-se de um efeito de *bottleneck* (efeito de gargalo), o primeiro dos muitos que se iriam observar no percurso de dispersão dos ciganos (Kalaydjieva *et al.*, 2005). Cada um foi certamente factor de perda de diversidade genética consequente à drástica redução do efectivo populacional (Jobling *et al.*, 2004),

A chegada dos ciganos a território persa é considerada a primeira entrada em território cristão (Fraser, 1998), e a sua permanência nessa região é mais reforçada pela existência de lendas, como a que relata que o monarca persa Bahram Gur persuadiu o rei da Índia a enviar-lhe músicos, que foram distribuídos pelas várias partes do reino da Pérsia, onde se estabeleceram e multiplicaram (Fraser, 1998). A maioria dos historiadores considera que a estadia na Pérsia foi de curta duração, com base em provas de linguística comparada. A conquista desta região pelos árabes ocorreu no ano de 646, o que provocou não só a mudança de religião mas também da língua, pois o uso oficial do persa foi absorvido pela língua árabe (Fraser, 1998). A ausência de vocábulos árabes no *Romani*,

sugere então, que a permanência dos ciganos na Pérsia precedeu a arabização da língua, ou não se prolongou muito após essa fase (Fraser, 1998).

Segundo alguns historiadores a passagem dos ciganos da Pérsia para a Arménia ocorreu durante o século VII, mas outros situam essa movimentação em meados do século XI, sob o efeito das incursões gregas na região (Fraser, 1998). Como consequência, os ciganos foram arrastados para o território de Bizâncio (Constantinopla e Trácia), que é considerado a grande “porta de entrada” da migração cigana em direcção à Europa Ocidental (Fraser, 1998; Kenderick e Puxon, 1998). A conquista dos gregos, nomeadamente de Antioquia, impôs o domínio bizantino numa região que se estendia da Ásia à Europa Oriental, o que facilitou a migração dos ciganos para o interior da Europa, atravessando o Mediterrâneo e provavelmente seguindo a rota dos peregrinos cristãos (Fraser, 1998).

A entrada dos ciganos no Sul da região dos Balcãs terá ocorrido durante os séculos XI e XII, e só depois se dispersaram pelo resto da Europa (Marushiakova e Popov, 1997; Fraser, 1998).

A exposição dos ciganos à influência grega deve ter sido intensa e prolongada, como indica a introdução de um maior número de vocábulos gregos do que persas no *Romani* (Fraser, 1998). Registos bem documentados da presença dos ciganos na Grécia datam de 1322, e foram identificados em Creta. Na segunda metade do século XIV, chegaram à ilha de Corfu, e muitos estabeleceram-se na região do Eiro, uma zona conhecida por “Pequeno Egipto” dada a grande fertilidade dos terrenos (Reyniers, 1995). Provavelmente, foi com base nesta designação da região que se enraizou no conhecimento popular a ideia que os ciganos tinham origem egípcia, ideia que perdurou durante muito tempo. O número de ciganos em Corfu era tão numeroso que se constituíram num feudo independente, o *feudum acinganorum*, que sobreviveu até ao século XIX (Fraser, 1998).

Da Grécia, os ciganos avançam para Norte e no século XV já se encontram espalhados por todo o continente europeu (Liégeois, 1989; Fraser, 1998).

É admitida a entrada dos ciganos na Península Ibérica, ainda durante o século XV, através da rota dos Pirinéus, como atestam os documentos mais antigos que dão conta da presença de ciganos no Norte da Península, e que são um salvo-conduto concedido por Afonso V de Aragão a “D. João do Egipto Menor”, datado de 12 de Janeiro de 1425, e outro concedido pouco depois ao “Conde Tomás do Pequeno Egipto”. Há ainda vagas

indicações de que a sua incursão na Península pode ter envolvido a passagem através das ilhas do Mediterrâneo ou via Gibraltar (Bataillard, *in* Costa, 2003) e até se tem especulado que o Norte de África terá sido a principal fonte dos ciganos ibéricos, tendo acompanhado árabes e mouros durante as invasões da Península Ibérica a partir de 771 (Costa, 2003), mas tratam-se de hipóteses pouco fundamentadas.

Após se estabelecerem na Península Ibérica, foi tão rápida a taxa de introdução de vocábulos espanhóis na língua original que quase acabaram por a perder. De tal forma que o *Caló*, a língua actualmente falada por alguns grupos gitanos de Espanha, é basicamente a mistura gramatical do espanhol, com retenção de muitos termos *Romani* (Fraser, 1998).

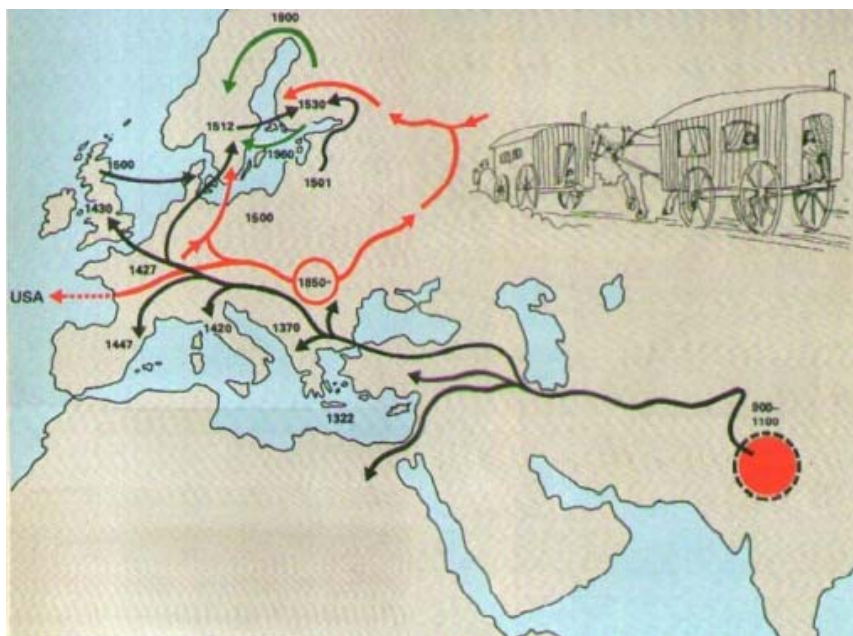


Figura 1. A dispersão das populações ciganas a partir da Índia em direcção à Europa Ocidental (imagem *in* www.vurdon.it/itinerario.jpg).

Uma segunda diáspora massiva de ciganos registou-se já no século XIX, quando os ciganos da Roménia perderam a condição de escravos. Essa migração envolveu a passagem de numerosos Roma para território americano, o que ocorreu paralelamente à sua dispersão pelo continente europeu. Contudo, a presença de ciganos na América não data dessa altura. Sabe-se que Cristóvão Colombo, na sua terceira viagem, em 1498, levou

os primeiros três ciganos que pisaram o “Novo Mundo”. Mais tarde, no século XVII, também a Inglaterra e a Escócia enviaram remessas de ciganos para as suas colónias americanas da Virgínia e Luisiana. A onda migratória diminuiu com o começo da Primeira Guerra Mundial, mas reiniciou-se em 1989 (Fraser, 1998).

A última grande diáspora cigana que decorreu mais recentemente com foco na região dos Balcãs, em direcção à Europa Ocidental, e considerada por muitos a terceira grande diáspora cigana, foi faseada em 2 ondas migratórias: a) a fragmentação da Jugoslávia durante os anos 70 a 90, impulsiona nova vaga de dispersão, particularmente para a Itália e Áustria, e é responsável pela sua presença nas grandes cidades da Europa Ocidental (Fraser, 1998) e b) finalmente, nas últimas duas décadas, registou-se a saída substancial de ciganos da Europa de Leste, novamente devido aos conflitos políticos e económicos dos países da região (Reyniers, 1995), tratando-se de uma movimentação ainda em curso.

Um aspecto em que convergem os estudos linguísticos, históricos, demográficos, antropológicos, culturais e mesmo genéticos, é na indicação de que quando se fala em ciganos não se pode pensar numa vaga de migração generalizada, mas antes em sucessivas migrações protagonizadas por grupos independentes mas muito coesos. Aparentemente, houve uma certa unidade de acção entre os grupos, que mantiveram uma ligação sólida entre si, como por vezes até sugere a semelhança das suas histórias, muito fantasiosas, que diferentes grupos Roma contam sobre a própria origem (Fraser, 1998).

Por outras palavras, os ciganos foram capazes de reconhecer na diversidade uma unidade interna, talvez baseada nas experiências comuns de desenraizamento, seja pela força física, seja pela força simbólica. Também importante, tais experiências de deslocamento puderam conectar-se com uma estrutura de parentesco semelhante (patrilinearidade e patrilocalidade) e uma política étnica de segmentação estrutural e espacial (Fazito, 2000).

1.1.2. Os ciganos portugueses

Existe grande dificuldade em descrever a comunidade cigana de Portugal. Ao longo do tempo, os elementos de identificação étnica foram sendo diluídos pelo abrandamento das estratégias endogâmicas que caracterizam amiúde as minorias étnicas (Costa, 2003), mas na realidade poucos estudos há incidindo sobre os ciganos portugueses.

Escasseiam as fontes históricas, as respectivas tentativas de compilação de vários estudos, e elementos coerentes que facultem quer comparações com outros Roma, quer o acompanhamento do que tem sido a permanência cigana no país. É de estranhar este pouco conhecimento sobre os ciganos portugueses, tanto mais que a sua presença secular em Portugal, que remonta já ao século XV, faz dos ciganos a comunidade mais antiga que, nos tempos modernos, adoptou Portugal como local de permanência (Costa, 2003). Também surpreende que um dos trabalhos de referência mais exaustivo sobre os ciganos portugueses continue a ser o publicado por Adolfo Coelho em 1892, alvo de recente reedição em 1995.

Sobre a chegada dos ciganos a Portugal, se Adolfo Coelho, em finais do século XIX, não encontra elementos que permitam garantir quando entraram os primeiros grupos em Portugal, volvido mais de um século a escassez de fontes mantém-se. Mas a maioria dos autores situa a chegada dos ciganos ao país em finais do século XV. De facto, logo no início do século seguinte, os ciganos são expressamente referidos em obras literárias como “*El Estudioso Cortesano*” de Lorencio Palmireno (1517) ou “*O Auto das Ciganas*” de Gil Vicente (1521), ainda que o documento mais antigo que refere os ciganos em terras portuguesas seja uma poesia da autoria de Luís da Silveira, intitulada “*As Martas de D. Jerónimo*”, publicada no Cancioneiro Geral de Garcia de Resende em 1510 (Coelho, 1995; Costa, 2003).

Não restam dúvidas que foi logo no século XVI, pouco depois de terem chegado a Portugal, que começaram as perseguições aos ciganos. Datam desse século os primeiros editos oficiais contra os ciganos, nomeadamente um decreto de 1538 que ordena a expulsão dos ciganos do país, após tortura física, e outro de 1598 que estipula a pena de morte para aqueles que não abandonassem o território nacional (Coelho, 1995). Os degredos, nomeadamente para o Brasil, iniciam-se ainda no século XVI e a prática de deportação intensifica-se nos dois séculos seguintes.

Só no início do século XIX foi conferido aos ciganos o estatuto de cidadãos portugueses, mas desde então os problemas de integração persistiram. Ainda hoje enfrentam sérios problemas de exclusão social, relativamente aos quais não se tem conseguido encontrar respostas adequadas.

Estima-se em cerca de 40 a 50 mil o número de ciganos actualmente estabelecidos em Portugal (Liégeois 1994, *in* Costa, 2003; fonte *www.coe.int*).

Ao contrário de outras populações ciganas, os ciganos de Portugal, tal como os gitanos de Espanha, não falam o *Romani* (Fraser, 1998). O “rumanho” (adaptação de rumano, romanó ou romano) é o termo usado pelos ciganos portugueses para designar a sua língua. Segundo Adolfo Coelho, no final do século XIX correspondia ao “*Caló* espanhol influenciado pelo português”, e a afinidade linguística indicava que os ciganos de Portugal eram um subgrupo dos gitanos de Espanha.

Hoje em dia, embora supostamente a língua continue a funcionar como um importante elemento de identidade cultural, para a maioria dos ciganos portugueses o rumanho é apenas uma referência linguística remota, pois falam basicamente o português com alguns termos remanescentes do rumanho.

Como se referiu no início deste ponto, poucos estudos existem sobre os ciganos portugueses.

O projecto em que se insere este trabalho, recorre a uma abordagem genética para estudar os ciganos portugueses, tendo em vista compreender como o seu padrão de diversidade se contextualiza no conjunto de outras populações ciganas e não ciganas, procurando, desta forma, recuperar aspectos da sua história populacional.

Os resultados já publicados (Gusmão *et al.*, 2008) e os agora apresentados neste trabalho, representam as primeiras caracterizações globais dos ciganos portugueses que revelam um padrão de diversidade muito peculiar, à semelhança do que se tem observado noutros grupos de ciganos da Europa. Trata-se de um padrão de diversidade que levou a considerar os ciganos da Europa como um “isolado genético transnacional”, segundo expressão de Kalaydjieva *et al.* (2005). É uma descrição que se compreende face à extraordinária dispersão geográfica dos grupos Roma, à tendência para um certo isolamento em grande parte devido às práticas sociais endogâmicas e, apesar de tudo, pela considerável coerência genética dos grupos. É um paradoxo que se explica pela partilha de uma origem recente comum.

Mas a par do fundo genético comum que os une, tem-se detectado em todos os ciganos da Europa proporções substanciais de mistura populacional, o que leva a relativizar a noção de cigano como grupo populacional fechado.

No fundo, é este conjunto de factores contraditórios nos efeitos, que faz dos ciganos um objecto de estudo especialmente interessante do ponto de vista da Genética Populacional.

1.2. Estudos em genética populacional e marcadores genéticos

Os estudos genéticos têm sido muito utilizados com o intuito de, conjuntamente com documentos e dados históricos, antropológicos, linguísticos, tentar esclarecer muitas dúvidas que persistem sobre a história das populações humanas. Na realidade, as achegas de investigação mais convencionais não têm sido suficientes para obter um quadro satisfatório da origem e evolução das populações humanas contemporâneas. A Genética tem proporcionado ferramentas de investigação com potencialidades únicas, em particular na tentativa de reconstruir os acontecimentos demográficos passados das populações (Jobling *et al.*, 2004). Não são raras as vezes em que os resultados da análise genética levantam novas questões que têm mobilizado uma atenção multidisciplinar.

O primeiro grande sucesso da genética populacional residiu em explicar como se podem prever numa população, em que os cruzamentos são ao acaso, proporções genotípicas numa dada geração com base nas frequências alélicas da geração anterior. O princípio foi formulado em 1908 e ficou conhecido como o princípio fundamental da genética populacional: princípio de Hardy-Weinberg. Esta área desenvolveu-se muito a partir de 1930 com os trabalhos realizados por Ronald Fisher e Sewall Wright (Jobling *et al.*, 2004). Mas desde então a natureza dos dados genéticos usados nos estudos populacionais foi sofrendo grandes alterações, em especial ao longo dos últimos 30 anos.

Começou por se recorrer aos chamados “marcadores clássicos”, nos quais se incluem os grupos sanguíneos e os marcadores protéicos ou enzimáticos. Na altura como se desconhecia a base molecular dos polimorfismos, as inferências evolutivas baseavam-se apenas na análise das frequências alélicas (Jobling *et al.*, 2004). Também o facto de apresentarem baixos níveis de variação genética e de muitos serem influenciados por

processos de selecção natural, representavam outras das limitações deste tipo de marcadores. Só mais tarde, com o desenvolvimento de novas técnicas da Biologia Molecular, foi possível analisar directamente o DNA e transferir o interesse dos marcadores clássicos para uma nova geração de marcadores, os “marcadores moleculares” (Jobling *et al.*, 2004). Muitos são altamente polimórficos, selectivamente neutros e facilmente genotipados por técnicas moleculares rápidas e simples baseadas na *Polimerase Chain Reaction* (PCR). Dentro dos marcadores moleculares, os RFLPs (*Restriction Fragment Length Polymorphisms*) foram os primeiros a serem usados em estudos de identificação humana e populacionais (Jeffreys *et al.*, 1985). Nas duas últimas décadas, o desenvolvimento e generalização de novas tecnologias permitiu que a atenção se voltasse para um conjunto de marcadores ainda mais informativos, entre os quais se incluem os microssatélites ou STRs (*Short Tandem Repeats*), que foram alvo de estudo neste trabalho.

Se a disponibilidade de marcadores genéticos proporcionou grandes avanços em estudos populacionais, não fez desaparecer muitas das dificuldades que surgem quando se procura alcançar uma compreensão refinada da história evolutiva das populações humanas. Em parte as dificuldades resultam do facto da teia de relacionamento entre populações humanas assentar em padrões filogenéticos bastante complexos que não dependem apenas da distância geográfica. Muitos aspectos do passado demográfico das populações, como migrações (que tendem a homogeneizá-las), ou alterações drásticas do efectivo populacional, incluindo *bottleneck* ou expansões populacionais, influenciam os padrões actuais de diversidade genética. Em especial tratando-se de populações com reduzidos efectivos populacionais, os efeitos estocásticos da deriva ou fluxo genético podem gerar intrincados padrões de subestruturação populacional.

Quando se procura inferir esses padrões com base em marcadores genéticos, as respostas podem diferir consoante o tipo de marcadores estudados. Assim apesar das histórias populacionais terem sido únicas, podem ser diferentes os sinais do passado que ficam retidos em distintos marcadores.

Compreende-se assim, que para reconstruir histórias populacionais seja importante recorrer a um número elevado e a diferentes tipos de marcadores genéticos.

No que respeita a populações ciganas, muitos dos trabalhos realizados foram baseados em marcadores clássicos (Bernasovsky *et al.*, 1976; Gyodi *et al.*, 1981 Sivakova *et al.*, 1983). Nos últimos anos, acompanhando naturalmente a tendência registada em

estudos similares, marcadores moleculares começaram a ser usados para estudar grupos Roma, e o grande interesse tem-se até centrado na análise do DNA mitocondrial (mtDNA) ou do cromossoma Y, que têm proporcionado informações importantes sobre a história dos ciganos (e.g. Kalaydjieva *et al.*, 2001b; Kalaydjieva e Morar, 2003; Malyarchuk *et al.*, 2006; Gusmão *et al.*, 2008).

No entanto, dadas as peculiares propriedades de transmissão, tanto o DNA mitocondrial como o cromossoma Y, fornecem visões parciais da história genética, nomeadamente das linhagens femininas (mtDNA) e das masculinas (cromossoma Y). Por outro lado, e uma vez que o efectivo populacional do DNA mitocondrial e do cromossoma Y é $\frac{1}{4}$ relativamente ao dos autossomas, ambos são muito susceptíveis a efeitos de deriva, e na medida em que constituem entidades genómicas não-recombinantes implica que cada um representa um único *locus*.

Importa, portanto, obter informações de mais marcadores e menos sujeitos a efeitos de deriva, razões pelas quais neste trabalho se tivesse optado por estudar STRs localizados nos autossomas e no cromossoma X.

1.2.1. Microssatélites ou *Short Tandem Repeats*

Os microssatélites ou STRs são polimorfismos em que os motivos de repetição são curtos, variando entre 2 a 4 nucleótidos, o que facilita bastante a sua análise e amplificação por PCR. São marcadores muito abundantes no genoma humano, estimando-se que ocorram em cada 10000 nucleótidos, e encontram-se dispersos por todos os cromossomas (autossómicos, cromossoma X e Y) (Budowle e Brown, 2001). São considerados marcadores de evolução rápida, visto exibirem uma taxa de mutação média bastante elevada 6×10^{-4} (Goldstein *et al.*, 1995).

Os STRs representam aproximadamente cerca de 3% do genoma humano, a grande maioria está presente nas regiões não-codificantes, sendo os dinucleótidos a classe mais comum nos humanos (Wren *et al.*, 2000).

As propriedades dos STRs são muito vantajosas em vários campos de aplicação, especificamente em estudos de genética populacional e forense, entre as quais se destacam: o elevado grau de polimorfismo, a sua dispersão uniforme pelo genoma humano,

possibilidade de genotipagem de forma simples e rápida, e reprodutibilidade dos resultados (Budowle e Brown, 2001).

Os microsátélites podem ser classificados de acordo com o motivo de repetição em “STRs simples”, se contêm um único motivo de repetição, e “STRs compostos”, se consistem em duas ou mais repetições simples adjacentes com diferentes motivos repetitivos (Butler, 2001). As características dos motivos repetitivos (tipo, número e proximidade) podem afectar a taxas de mutação e a variação de alelos dos STRs. Deste modo, *loci* mais extensos, isto é, com maior número de motivos de repetição, encontram-se associados a um nível de polimorfismo mais elevado do que os mais pequenos (Weber e Wong, 1993).

Um aspecto que se tem procurado elucidar diz respeito aos mecanismos responsáveis pela origem de mutações em STRs. Diversos estudos indicam que com o aumento do tamanho da estrutura de repetição, aumenta a probabilidade de ocorrerem mutações pontuais, dividindo o motivo inicial em outros estruturalmente mais pequenos e com menores taxas de mutação (Bell e Jurka, 1997; Kruglyak *et al.*, 1998; Calabrese *et al.*, 2001).

Em estudos evolutivos e de genética populacional, são muito importantes os modelos mutacionais que se assumem para a geração de variabilidade genética. Os dois modelos mais convencionais são o *Infinite Allele Mutation Model* (IAMM), de Kimura e Crow (1964) e o *Stepwise Mutation Model* (SMM), estabelecido por Kimura e Ohta (1973). Segundo o IAMM, cada mutação pode originar um alelo novo qualquer e assim afinidades moleculares entre alelos não são indicativas de relação filogenética mais próxima. No modelo SMM assume-se que novos alelos surgem por pequenos passos mutacionais e portanto as diferenças moleculares entre dois alelos são filogeneticamente informativas, ou seja, quanto maior a diferença maior o número de eventos mutacionais ocorrido desde a origem comum de dois alelos.

Continua a ser habitual assumir um dos dois modelos mais convencionais, IAMM ou SMM, e a opção por um ou outro modelo tem que ser bem ponderada quando se pretende inferir afinidades genéticas interpopulacionais. De facto, quando se usam dados de STRs para comparar populações pode recorrer-se ao clássico F_{ST} (Wright, 1951), que assume o modelo IAM, ou à distância genética R_{ST} (Slatkin, 1995), que assume o modelo SMM. Embora a segunda medida pareça mais apropriada para lidar com STRs, alguns

estudos têm indicado que a sua aplicação pode confundir a teia de relacionamento entre populações, sobretudo quando estão em causa populações em que efeitos de deriva tiveram um papel de relevo ao longo da história (Balloux e Goudet, 2002). Nestes casos, o clássico F_{ST} , ao dar menos peso ao processo mutacional, pode reflectir melhor os mecanismos e forças demográficas subjacentes ao relacionamento entre populações (Cagliá *et al.*, 2003), sendo o erro associado a este parâmetro menor relativamente ao R_{ST} , quando o efectivo populacional é razoável (Gaggioti *et al.*, 1999).

1.3. Cromossoma X e autossomas

O cromossoma X contém uma molécula de DNA linear com aproximadamente 165 Mb, representando cerca de 5% do genoma humano total. Uma característica singular deste cromossoma sexual é a sua baixa densidade génica comparativamente aos autossomas, compreendendo apenas 1100 genes dos 20000 a 25000 que se estima existir em todo o genoma humano (Venter *et al.*, 2001)

Os dois cromossomas sexuais, X e Y, tiveram origem a partir de um único par de autossomas homólogos que há cerca de 300 milhões de anos começaram a divergir (Ohno, 1967; Lahn e Page, 1999), mantendo a homologia e recombinação um com o outro nas regiões pseudoautosómicas. Apesar da origem comum, os cromossomas sexuais seguiram vias evolutivas diferentes (Ross *et al.*, 2005), pois enquanto o cromossoma Y perdeu grande parte do material genómico, incluindo os genes que partilhava com o cromossoma X (Skaletsky *et al.*, 2003; Jobling e Tyler-Smith, 2003), o cromossoma X não perdeu o seu carácter autossómico a nível da expressão, sendo até o cromossoma nuclear fisicamente mais estável nos mamíferos placentários (Waterson *et al.*, 2002).

As mulheres herdam dois cromossomas X de cada progenitor, mas os homens herdam um único cromossoma X transmitido por via materna. As descendentes femininas recebem um cromossoma X do pai de forma directa, ou seja, sem sofrer recombinação excepto nas regiões pseudoautosómicas. Pelo contrário, o cromossoma proveniente da mãe resulta de recombinação entre os seus dois cromossomas X, possuindo assim informação genética de ambos os cromossomas maternos (Schaffner, 2004; Ross *et al.*, 2005).

Em relação aos homens o único cromossoma que herdamos possui informação dos dois cromossomas X maternos em resultado do processo de recombinação. Como nos indivíduos de sexo masculino o X não tem homólogo, a sua análise permite a determinação directa de haplótipos (Schaffner, 2004; Ross *et al.*, 2005).

Para além das propriedades de transmissão, outros aspectos diferem entre cromossoma X e autossomas, conforme recentemente revisto por Schaffner (2004). Em termos de efectivo populacional, o número de cromossomas X equivale a $\frac{3}{4}$ do dos autossomas, o que implica no cromossoma X efeitos de deriva génica mais acentuados. Em consequência, observa-se uma maior estruturação populacional no cromossoma X que em autossomas.

Em média, são mais baixas as taxas de mutação do X comparativamente a autossomas, pois os cromossomas X permanecem $\frac{2}{3}$ da sua vida em mulheres, e nestas as taxas de mutação tendem a ser inferiores às dos homens (Li *et al.*, 2002).

Uma outra particularidade do cromossoma X é que neste cromossoma o nível de desequilíbrio de ligação (LD) é mais alto e estende-se por intervalos mais longos que em autossomas, na medida em que a taxa de recombinação no X corresponde apenas a cerca de $\frac{2}{3}$ da taxa de recombinação genómica. Tal facto resulta do cromossoma X só recombinar nas mulheres (Schaffner, 2004). A magnitude de LD no cromossoma X é especialmente elevada em populações de tamanho reduzido e geneticamente isoladas (Zavattari *et al.*, 2000; Katoh *et al.*, 2002), assim como em populações miscigenadas, onde também têm sido observados fortes níveis de LD (Wilson e Goldstein, 2000).

1.3.1. STRs do cromossoma X e autossómicos

Desde há muito que os STRs autossómicos têm vindo a ser utilizados em estudos de genética populacional e em aplicações forenses, sendo vastíssimo o catálogo conhecido de STRs autossómicos.

Neste trabalho optou-se por analisar um conjunto de 15 STRs que foi optimizado tendo em vista a utilização primária na área forense. Por essa razão, têm vindo a ser publicados muitos dados populacionais, incluindo alguns sobre populações ciganas, ainda

que não abundem dados contemplando o conjunto completo de STRs autossómicos aqui abordados.

Já quanto ao potencial dos polimorfismos do cromossoma X, só nos últimos anos é que começou a ser explorado, bastante depois da explosão de interesse pelo cromossoma Y e DNA mitocondrial, quer em estudos de populações quer em análises forenses. Deste modo, e também pelo facto de se tratar de um único cromossoma, o número de polimorfismos do cromossoma X descritos é ainda muito limitado. No entanto os STRs do X têm revelado ser ferramentas valiosas em testes de parentesco, principalmente em casos complexos de paternidade em que a criança em disputa é do sexo feminino (Edelmann *et al.*, 2002b), e complementos de análise importantes em antropologia evolutiva (Szibor *et al.*, 2003b; Szibor *et al.*, 2006).

Tal como para os STRs autossómicos, também os X-STRs estudados neste trabalho têm sido muito explorados em investigações forenses. Quanto aos STRs mais interessantes nesse contexto, estudos diversos permitiram identificar quatro grupos de ligação (Szibor *et al.*, 2003a), conforme se apresenta na figura 2.

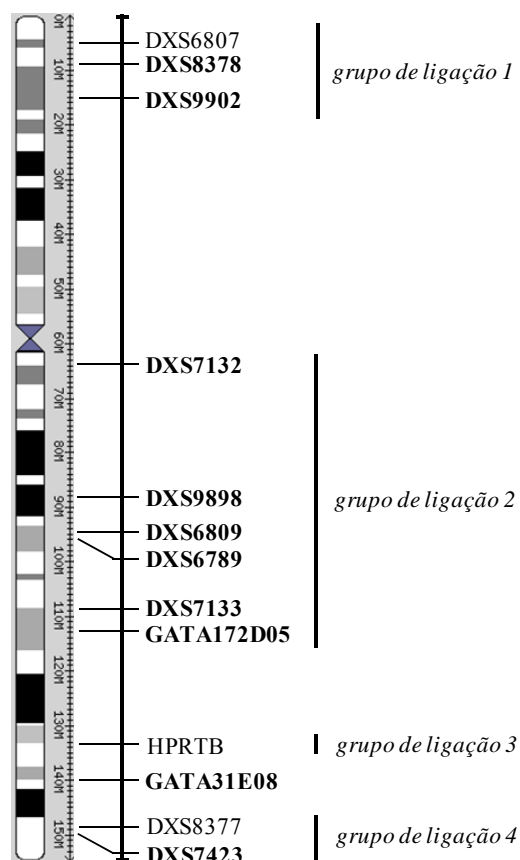


Figura 2. Localização dos 10 STRs do cromossoma X estudados neste trabalho (a **negrito**) e respectivos grupos de ligação com indicação dos STRs flanqueantes, de acordo com Szibor *et al.* (2003a).

1.4. Objectivos

No sentido de aprofundar o conhecimento sobre a origem e história demográfica dos ciganos portugueses, neste trabalho caracterizou-se uma amostra deste grupo populacional com o intuito de complementar o estudo recentemente efectuado por Gusmão *et al.* (2008). As investigações desenvolvidas por Gusmão e colaboradores, foram pioneiras na área da genética populacional sobre os ciganos de Portugal, tendo incidido sobre a análise dos padrões de diversidade genética ao nível de marcadores do cromossoma Y.

A mesma amostra de ciganos portugueses não-aparentados, foi agora analisada para dois conjuntos diferentes de marcadores genéticos, nomeadamente 15 microssatélites autossómicos e 10 localizados no cromossoma X, com os seguintes objectivos:

- aferir as similaridades genéticas entre ciganos portugueses e outros grupos Roma;
- obter indicações sobre o grau de fluxo génico entre os ciganos portugueses e outras populações ciganas e não-ciganas;
- avaliar as afinidades genéticas dos grupos Roma, em particular dos ciganos portugueses, com potenciais populações parentais;
- comparar estimativas de mistura populacional nos ciganos portugueses com as de outros grupos Roma;
- averiguar que sinais do passado demográfico ficaram retidos no padrão de diversidade dos grupos ciganos actuais;
- contribuir para um melhor conhecimento da história das populações ciganas europeias.

2. Material e Métodos

2.1. Amostragem e material biológico

A amostra analisada neste trabalho foi recolhida em 18 comunidades ciganas diferentes dispersas por 11 distritos de Portugal. O total da amostra compreende 127 indivíduos não aparentados, 123 homens e 4 mulheres, com naturalidade em 16 distritos do país: Aveiro ($N=8$); Beja ($N=8$); Braga ($N=12$); Bragança ($N=12$); Castelo Branco ($N=5$); Coimbra ($N=4$); Évora ($N=1$); Faro ($N=6$); Funchal ($N=1$); Leiria ($N=5$); Lisboa ($N=42$); Portalegre ($N=1$); Porto ($N=13$); Santarém ($N=2$); Setúbal ($N=5$) e Vila Real ($N=2$) (figura 3).

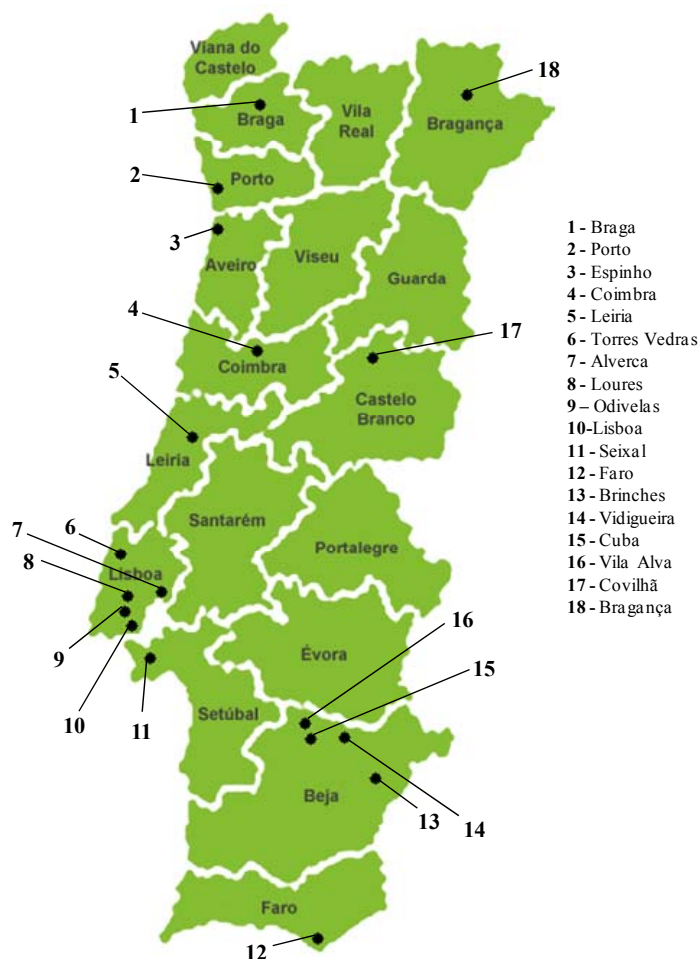


Figura 3. Mapa de Portugal indicando a localização das zonas de colheita das amostras dos 127 indivíduos da população cigana portuguesa.

O material biológico usado foi sangue obtido por picada de dedo com lanceta *Accu-Check SafeT-Pro Plus* (Roche Diagnostics GmbH, Mannheim, Germany), com sistema de punção esterilizado de utilização única. Posteriormente foi conservado em cartões FTA (FTA Classic Cards – Whatman BioScience Ltd., Abington, Cambridge, UK), formando uma mancha com cerca de 1 cm². Os cartões FTA facilitam o processo de recolha e constituem um excelente suporte para armazenagem e conservação de amostras de material biológico. De facto, uma pequena porção de papel impregnado com sangue é suficiente para obter DNA genómico em quantidade adequada para proceder à respectiva purificação e amplificação. Por outro lado, é apenas necessário seguir regras básicas de manipulação para evitar problemas de contaminação entre amostras.

2.2. Extração de DNA

DNA genómico foi extraído recorrendo a um método convencional de fenol-clorofórmio seguindo os seguintes passos:

- Cortar cerca de 3mm² da mancha de sangue impregnada no cartão FTA em pedaços mais pequenos para um tubo e adicionar 575 µl de solução tampão DLB (10 ml TRIS/HCl 1M pH 7.4; 2 ml NaCl 5M; 20 ml EDTA 0.5M pH 8.0 e 968 ml H₂O);
- Adicionar 50 µl de SDS a 10% e 5 µl de proteinase K (20 mg/ml);
- Incubar a 56°C em banho-maria durante a noite (±12 horas) com agitação suave;
- Adicionar 20 µl de NaCl 5M e 575 µl de solução fenol-clorofórmio/álcool isoamílico (25:24:1) e agitar;
- Centrifugar durante 3 min a 12000 rpm;
- Recuperar o sobrenadante para outro tubo e descartar o remanescente;
- Adicionar 1 ml de etanol a 96% conservado a -20°C;
- Colocar a -80°C durante 25 min;
- Centrifugar a frio durante 15 min a 12000 rpm;
- Descartar o etanol, secar o precipitado e ressuspender com a adição de 100 µl de H₂O colocando num agitador a 36°C durante 12 horas;
- Armazenar o DNA à temperatura de -20°C até à sua utilização.

2.3. Análise de microssatélites autossómicos e do cromossoma X

2.3.1. Amplificação por PCR multiplex

Na amplificação dos STRs autossómicos foram usados *primers* marcados com cinco fluorocromos diferentes, o que permitiu com uma única reacção amplificar e detectar os produtos de 15 *loci* distintos estudados neste trabalho. O multiplex usado foi o AmpF/STR Identifiler, que inclui no total os 15 marcadores e o teste de determinação do sexo, o *locus* da amelogenina (A). Dos 15 STRs, 13 estão incluídos no *Combined DNA Index System* (CODIS) (Budowle *et al.*, 1998): D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, vWA, TPOX, D18S51, D5S818 e FGA, e os dois STRs adicionais que também são internacionalmente aceites: D2S1338 e D19S433 (Collins *et al.*, 2004) (figura 4).

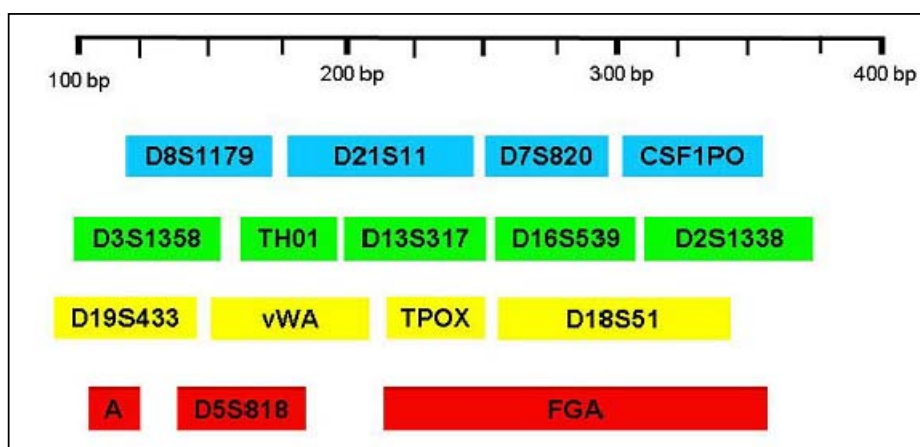


Figura 4. Diagrama esquemático do multiplex usado para amplificar 15 STRs autossómicos, onde se apresentam as cores atribuídas a cada microssatélite e a gama de variação do tamanho esperado, em pares de bases (bp), para os alelos de cada STR.

Na tabela 1 encontram-se descritos os motivos de repetição de cada marcador autossómico e a sua referência no *GeneBank*, que tem em conta a respectiva localização na

sequência do genoma humano, de acordo com Butler *et al.* (2006), o que resulta da colecção de dados obtidos em trabalhos anteriores.

Tabela 1. Características dos marcadores STR autossómicos estudados neste trabalho.

<i>Locus</i>	<i>Motivo repetitivo</i>	<i>Alelos</i>	<i>Ref. GeneBank</i>	<i>Referência</i>
D8S1179	[TCTA][TCTG]	7-20	AF216671	http://www.chlc.org (GATA7G07)
D21S11	Complexo [TCTA] [TCTG]	12-41.2	AP000433	Sharma e Litt, 1992
D7S820	[GATA]	5-16	AC004848	http://www.chlc.org (GATA3F01.511)
CSF1PO	[TAGA]	5-16	X14720	Hammond <i>et al.</i> , 1994
D3S1358	[TCTG][TCTA]	8-21	AC099539	Li <i>et al.</i> , 1993
TH01	[TCAT]	3-14	D00269	Polymeropoulos <i>et al.</i> , 1991
D13S317	[TATC]	5-16	AL353628	http://www.chlc.org (GATA7G10.415)
D16S539	[GATA]	5-16	AC024591	http://www.chlc.org (GATA11C06)
vWA	[TCTG][TCTA]	10-25	M25858	Kimpton <i>et al.</i> , 1992
TPOX	[GAAT]	4-16	M68651	Anker <i>et al.</i> , 1992
D18S51	[AGAA]	7-39.2	AP001534	Straub <i>et al.</i> , 1993
D5S818	[AGAT]	7-18	AC008512	http://www.chlc.org (GATA3F03.512)
FGA	[CTTT]	12.2-51.2	M64982	Mills <i>et al.</i> , 1992
D2S1338	[TGCC][TTCC]	15-28	AC010136	http://www.chlc.org (GGAA3A09)
D19S433	AAGG	9-17.2	AC008507	http://www.chlc.org (GGAA2A03)

Para as reacções de amplificação seguiram-se as instruções recomendadas no kit AmpFI STR Identifiler (AB Applied Biosystems, Foster city, CA). Usou-se um volume de reacção final de 10 µl, com a seguinte composição: 3,8 µl de PCR *master mix*; 2 µl de mistura de *primers*, cada um com uma concentração de 0,2 µM; 0,2 µl da enzima DNA polimerase; 3 µl de água e 1 µl de DNA genómico com uma concentração entre 0,5-1,25 ng/ µl. As condições de PCR consistiram numa desnaturação inicial de 15 minutos a 95°C, seguida de 28 ciclos durante 1 minuto a 94°C, 1 minuto a 59°C e 1 minuto 72°C, e finalmente um passo de extensão final de 60 minutos a 60°C. As reacções de amplificação foram todas realizadas em termocicladores GeneAmp 2700 (AB Applied Biosystems).

Em cada reacção de amplificação foi usada uma amostra como controlo negativo, onde o DNA foi substituído por água estéril, e uma amostra como controlo positivo contendo DNA de genótipo conhecido, nomeadamente a amostra de referência 9947A (AmpFI STR Identifiler PCR Amplification Kit; AB Applied Biosystems) (tabela 2).

Tabela 2. Perfil de STRs autossómicos da amostra de DNA (9947A) usada como controlo positivo descrito no manual do *kit* de PCR AmpF/STR Identifier (*AB Applied Biosystems*).

<i>Locus</i>	<i>Alelos</i>
D8S1179	13
D21S11	30
D7S820	10-11
CSF1PO	10-11
D3S1358	14-15
TH01	8-9.3
D13S317	11
D16S539	11-12
vWA	17-18
TPOX	8
D18S51	15-19
D5S818	11
FGA	23-24
D2S1338	19-23
D19S433	14-15

Relativamente aos marcadores do cromossoma X foram analisadas apenas as amostras pertencentes aos indivíduos do sexo masculino ($N=123$). O processo de amplificação foi também realizado em sistema multiplex, usando o *kit* Quiagen®Multiplex PCR (QIAGEN; www.qiagen.com), o que permitiu a obtenção de f

ragmentos para cada um dos 10 marcadores estudados (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902 e DXS6789) (figura 5). A utilidade deste sistema multiplex consiste na possibilidade de amplificação simultânea dos marcadores citados, numa única reacção de PCR.

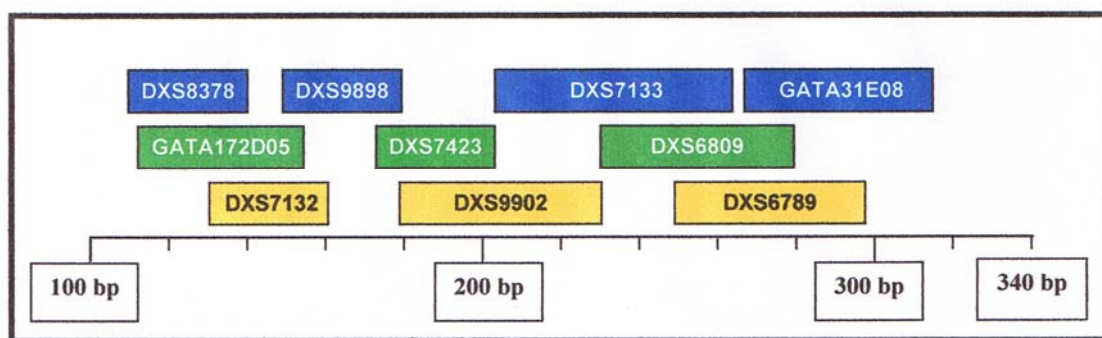


Figura 5. Representação esquemática do multiplex usado para genotipar os 10 X-STRs, com referência aos intervalos de tamanho esperados para os alelos de cada STR, em pares de bases (bp), e a marcação fluorescente usada em cada microsatélite (6-FAM, azul; VIC, verde e, NED, amarelo).

O *set de primers* escolhido para utilizar neste trabalho foi o descrito por Gusmão *et al.* (2009), e estão apresentados na tabela 3, assim como as respectivas marcações com fluorocromos.

Tabela 3. Sequências dos primers usados e respectiva marcação com fluorocromos.

Marcador	Fluorocromo	Sequências dos <i>primers</i> (5'-3')	Referência bibliográfica
DXS8378	6-FAM	TTAGGCAACCCGGTGGTCC ACAAGAACGAAACTCCAAC	Edelmann <i>et al.</i> , 2002a
DXS9898	6-FAM	CGAGCACACCTACAAAAGCTG TAGGCTCACCTCACTGAGCA	Gomes <i>et al.</i> , 2007
DXS7133	6-FAM	CACTTCCAAAAGGGGAAAAA ACTTGTAAGTGGTGGGAGGAA	Gusmão <i>et al.</i> , 2009a
GATA31E08	6-FAM	GCAAGGGGAGAAGGCTAGAA TCAGCTGACAGAGCACAGAGA	Gusmão <i>et al.</i> , 2009a
GATA172D05	VIC	TAGTGGTGATGGTTGCACAG ATAATTGAAAGCCCGGATTC	Edelmann <i>et al.</i> , 2002a
DXS7423	VIC	GTCTTCCTGTCATCTCCCAAC TAGCTTAGCGCCTGGCACATA	Edelmann <i>et al.</i> , 2002a
DXS6809	VIC	TCCATCTTTCTCTGAACCTTCC TGCTTTAGGCTGATGTGAGG	Gomes <i>et al.</i> , 2007
DXS7132	NED	TCCCCTCTCATCTATCTGACTG CACTCCTGGTGCCAAACTCT	Gomes <i>et al.</i> , 2007
DXS9902	NED	CTGGGTGAAGAGAAGCAGGA GGCAATACACATTCATATCAGGA	Gusmão <i>et al.</i> , 2009a
DXS6789	NED	CTTCATTATGTGCTGGGGTAAA ACCTCGTGATCATGTAAGTTGG	Gomes <i>et al.</i> , 2007

A nomenclatura dos *loci* do cromossoma X tem por base o número total de repetições do motivo repetitivo.

Na tabela 4 apresenta-se a nomenclatura, o motivo de repetição e os alelos descritos para cada X-STR (à semelhança do processo seguido para os STRs autossómicos), resultado da compilação de vários estudos recentemente efectuados.

Tabela 4. Características dos marcadores X-STR incluídos no sistema multiplex utilizado neste trabalho.

<i>Locus</i>	Motivo repetitivo	Alelos	Referência
DXS8378	[CTAT] _n	8-14	Edelmann <i>et al.</i> , 2001
DXS9898	[TATC] ₂ ATC [TATC] _n [ATC] ₀₋₁ [TATC] _n	8.3-15	Hering e Szibor, 2000
DXS7133	[ATAG] _n	7-13	Edelmann <i>et al.</i> , 2001
GATA31E08	[AGGG] ₂₋₃ [AGAT] _n	7-15	Gomes <i>et al.</i> , 2009
GATA172D05	[TAGA] _n	6-12	Edelmann <i>et al.</i> , 2002a
DXS7423	[TCCA] _n	14-17	Zarrabeitia <i>et al.</i> , 2002
DXS6809	[CTAT] _n [ATCT] ₃ N₉ [TACT] ₃₋₄ [ATCT] ₃₋₅ N₁₀ [ATCT] _n	27-38	Edelmann <i>et al.</i> , 2003
DXS7132	[CTAT] _n	11-17	Edelmann <i>et al.</i> , 2001
DXS9902	[TAGA] _n	9-13	Gusmão <i>et al.</i> , 2009
DXS6789	[TATC] ₀₋₁ [TATG] _n	14-25	Hering <i>et al.</i> , 2001

O protocolo de amplificação por PCR seguido foi de acordo com Gusmão *et al.* (2009), utilizando o *kit* QIAGEN Multiplex (<http://www.qiagen.com>). Usou-se um volume final igual a 5 µl composto por: 2,5 µl de QIAGEN multiplex *master mix*; 0,5 µl de *primer mix* a uma concentração de 2 µM; 1 µl de água estéril e 1 µl de DNA genómico. As condições cíclicas de PCR tiveram início com uma desnaturação de 15 minutos a 95°C, seguida de 10 ciclos de 30 segundos a 94°C, 1 minuto a 60°C e 40 segundos a 72°C; 20 ciclos de 30 segundos a 94°C, 1 minuto a 58°C e 40 segundos a 72°C. Houve alteração dos tempos de *annealing* e de extensão dos respectivos *primers*, relativamente aos descritos por Gusmão *et al.* (2009), com o intuito de otimizar o processo de amplificação para o decaplex. Por último foi realizado um passo de extensão durante 60 minutos a 72°C. O processo de amplificação realizou-se em termocicladores GeneAmp 2700 (*AB Applied Biosystems*).

Tal como para os marcadores autossómicos, na amplificação dos X-STRs também foi usado um controlo negativo, substituindo uma amostra de DNA por água estéril, e um

controlo positivo que foi a amostra de referência 9947A, cujo perfil genotípico se apresenta na tabela 5. Quanto à nomenclatura dos alelos dos X-STRs seguiram-se as recomendações de Gusmão *et al.* (2009), que propõem para GATA31E08 e DXS9902 a adição de dois e um motivo de repetição, respectivamente, em relação ao inicialmente descrito por Shin *et al.* (2005) e Szibor *et al.* (2003a).

Tabela 5. Perfil de X-STRs da amostra 9947A usada como controlo positivo.

<i>Locus</i>	<i>Alelos</i>	<i>Referência</i>
DXS8378	10-11	Szibor <i>et al.</i> , 2003a
DXS9898	12-15	Szibor <i>et al.</i> , 2003a
DXS7133	9-10	Szibor <i>et al.</i> , 2003a
GATA31E08	13	Gomes <i>et al.</i> , 2009
GATA172D05	10	Szibor <i>et al.</i> , 2003a
DXS7423	14-15	Szibor <i>et al.</i> , 2003a
DXS6809	31-34	Szibor <i>et al.</i> , 2003a
DXS7132	12	Szibor <i>et al.</i> , 2003a
DXS9902	12	Gusmão <i>et al.</i> , 2009
DXS6789	21-22	Szibor <i>et al.</i> , 2003a

2.3.2. Genotipagem

Os produtos resultantes da amplificação por PCR multiplex (STRs autossómicos e X-STRs) foram separados por electroforese capilar em condições desnaturantes, utilizando o polímero POP-7 (*Performance Optimized Polymer 7*) (AB Applied Biosystems), e posteriormente foram detectados num sequenciador automático de 4 capilares ABI 3130 Genetic Analyser (AB Applied Biosystems). Para o efeito procedeu-se da seguinte forma: a cada 1 µl de amplificado foram adicionados 12 µl de uma mistura composta por formamida desionizada 95% - *Hi-Di Formamide* (AB Applied Biosystems) e marcador interno de peso molecular GeneScan-500 LIZ (AB Applied Biosystems), numa proporção de 33:1. Para a determinação do tamanho dos fragmentos amplificados foi utilizado o *software* GeneMapper 4.0 (AB Applied Biosystems).

Por último, a genotipagem foi realizada por comparação com o perfil alélico dos *ladders* incluídos em cada um dos *kits* de marcadores utilizados, STRs autossómicos (figura 6) e X-STRs (figura 7).

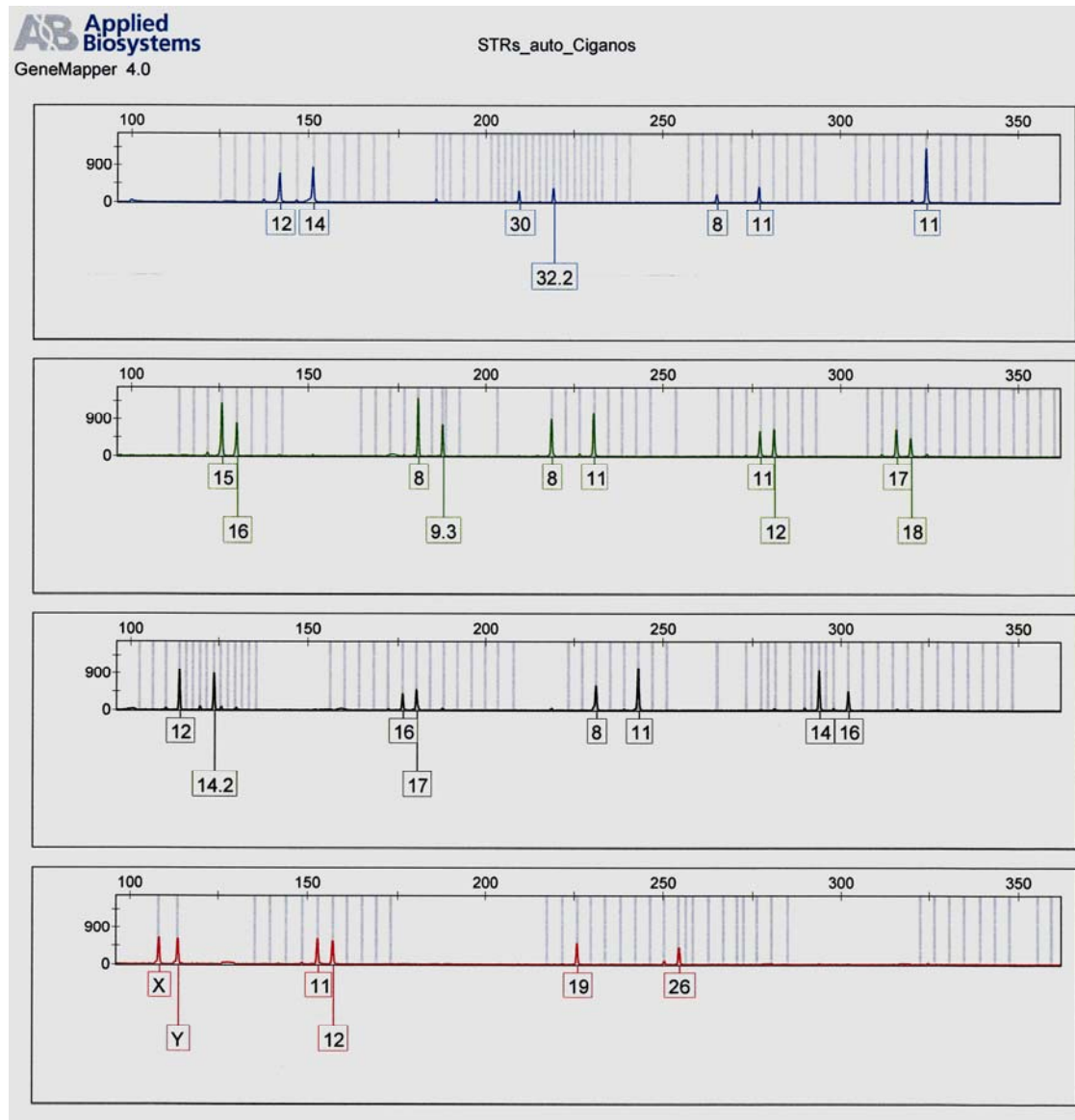


Figura 6. Electroforegrama do perfil de um indivíduo obtido com o *kit* AmpF/STR Identifier. **Nota:** a preto encontram-se representados os STRs correspondentes ao fluorocromo amarelo.

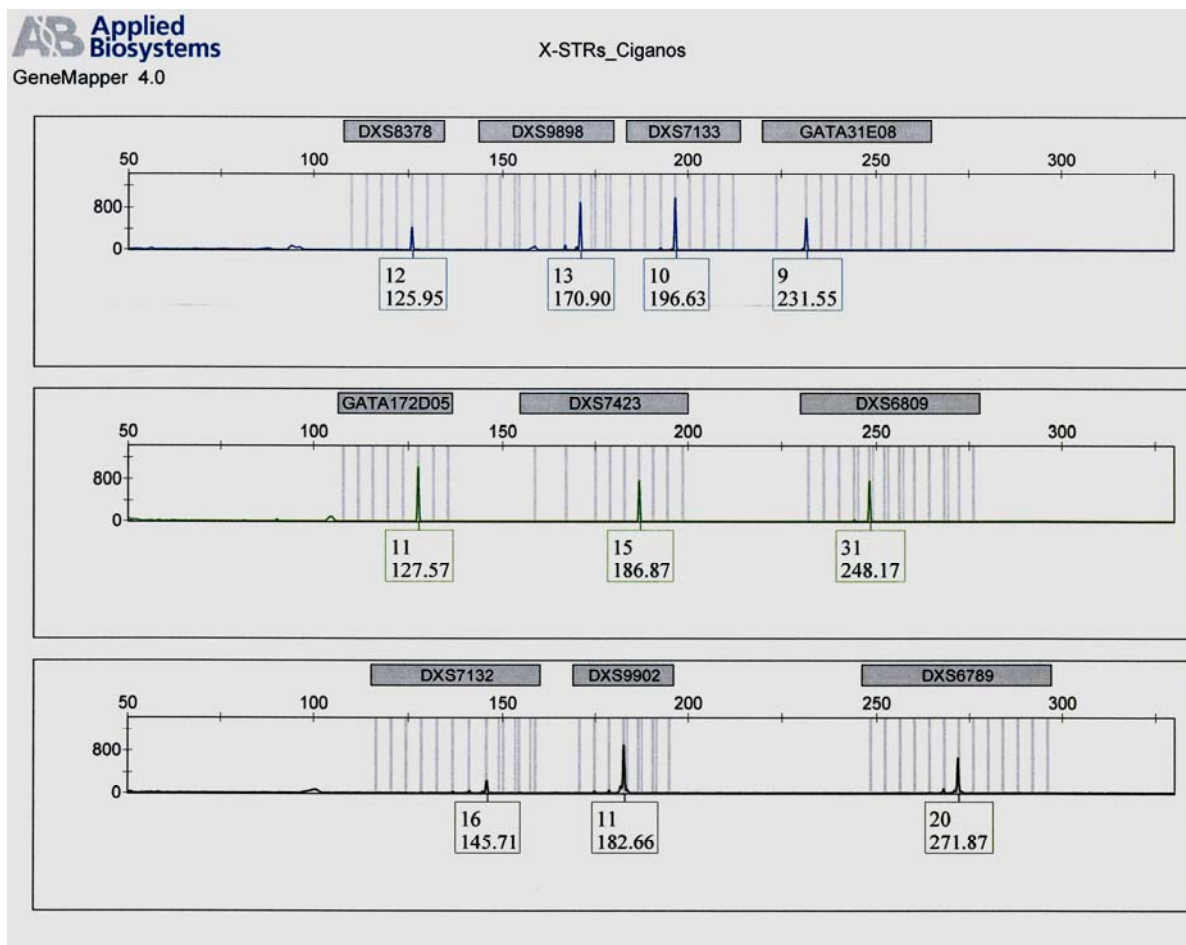


Figura 7. Electroforegrama do perfil de um indivíduo do sexo masculino obtido com o sistema X-decaplex.
Nota: a preto encontram-se representados os STRs correspondentes ao fluorocromo amarelo.

2.4. Análise de dados

Para descrever a diversidade genética da população cigana portuguesa, foram calculados diversos parâmetros descritivos e testou-se a significância dos desvios entre valores observados e previstos, assumindo o formalismo de Hardy-Weinberg.

Procedeu-se também a uma análise comparativa entre a população cigana portuguesa e outras populações ciganas e não-ciganas, após reunião de dados relevantes previamente publicados.

Discriminam-se de seguida os métodos usados para STRs autossómicos e do cromossoma X.

2.4.1. Análise de marcadores STRs autossómicos

As frequências alélicas para cada STR foram estimadas por contagem directa recorrendo ao programa Arlequin versão 3.01 (Excoffier *et al.* 2005). Este programa foi também utilizado para: a) obter valores de diversidade genética para cada marcador, b) testar o equilíbrio de Hardy-Weinberg para as distribuições genotípicas observadas, e c) calcular distâncias genéticas globais (F_{ST}) e *locus a locus*, entre populações e respectivos valores de significância estatística, efectuando 1000 permutações entre pares de populações. Considerou-se o nível de significância de $p=0,05$, e quando o mesmo conjunto de dados foi aplicado em múltiplos testes aplicou-se a correcção de Bonferroni, para estabelecer o nível de significância. Este factor de correcção foi determinado através do quociente entre o nível de significância adoptado (5%) e o número de marcadores ($N=15$), resultando num novo $p=0,003$.

Como o número de STRs comuns estudados em cada população era bastante variável, calcularam-se reiteradamente distâncias genéticas considerando 4, 7 ou 13 STRs partilhados.

Com base nos valores de F_{ST} entre populações obtidos com 13 *loci*, foi construída uma árvore filogenética pelo método de *Neighbour Joining*, integrado no programa Phylip

versão 3.65. A árvore foi posteriormente visualizada com o programa FigTree versão 1.2 (<http://tree.bio.ed.ac.uk/>).

De forma a obter uma representação gráfica do padrão de diferenciação interpopulacional aplicaram-se dois tipos de métodos: a) *Multidimensional Scalling* (MDS), com base nas referidas distâncias genéticas (F_{ST}), e b) Análise de Componentes Principais (PCA), com base nas frequências alélicas em cada marcador. Ambas as análises foram efectuadas através de rotinas apropriadas, incluídas no pacote de programa Statistica versão 8.0 (StatSoft, Inc. 2007; www.statsoft.com).

Por fim, foram estimadas taxas de miscigenação, quer na população cigana portuguesa quer noutras populações ciganas europeias (Eslováquia, Macedónia e Croácia). Admitindo a existência de três populações parentais, avaliou-se o grau de contribuição de cada uma delas na formação das populações ciganas. As três populações parentais consideradas foram: a) Europa, b) Turquia e Afeganistão em conjunto e c) Índia. Através do programa Admix versão 2.0 (Dupanloup e Bertorelle, 2001) foram calculados os coeficientes de mistura (mY_i), descritos por Bertorelle e Excoffier (1998), com base nos valores de frequências génicas assumidas para os grupos parentais. As proporções de mistura e respectivos erros resultaram de 10000 réplicas efectuadas ao acaso. A taxa de mutação por ano considerada foi de 0,000052, tendo em conta o valor médio da taxa de mutação dos 13 *loci* publicada individualmente no *Annual Report of the American Association of Blood Banks* de 2003 (www.aabb.org), e assumindo um tempo médio de geração de 25 anos.

2.4.2. Análise de marcadores X-STRs

Os valores das frequências alélicas e diversidade genética para cada marcador foram obtidos de forma idêntica à descrita para os microsátélites autossómicos, recorrendo novamente ao programa Arlequin versão 3.01 (Excoffier *et al.*, 2005). Como só se estudaram indivíduos do sexo masculino, estimaram-se pelo método de contagem directa as frequências haplotípicas considerando os 10 X-STRs. Os respectivos valores bem como os valores de diversidade haplotípica foram obtidos também com o Arlequin versão 3.01.

Com o intuito de detectar o grau de associação entre marcadores do cromossoma X na população cigana portuguesa, testou-se o desequilíbrio de ligação (*linkage disequilibrium* -LD), entre os 45 pares possíveis considerando os 10 X-STRs estudados.

A significância das associações foi obtida com o Arlequin versão 3.01, e como se realizaram múltiplos testes, aplicou-se novamente a correcção de Bonferroni. Assim, o quociente entre o nível de significância de 5% e o número de testes independentes realizados ($N=45$) definiu o valor de $p=0,001$, abaixo do qual se consideram associações estatisticamente significativas.

Foi também com o Arlequin que se obtiveram os valores de D' (valor de *linkage* entre pares de marcadores) e r^2 (associação global *multilocus*) entre pares de alelos de diferentes *loci*.

Como o decaimento de LD pode ser influenciado pela distância física entre *loci*, apresentam-se na tabela 6 as localizações físicas de cada um dos 10 STRs em estudo, obtidas por consulta da base de dados NCBI *build* 36 (Machado e Medina-Acosta, 2009), e a correspondente localização genética, definida em centiMorgans (cM), retirada do mapa do genoma humano *Rutgers* disponível online (<http://compugen.rutgers.edu/map-interpolar>).

Tabela 6. Marcadores STR descritos e sua localização física e genética no cromossoma X.

X-STR	Localização física (bp)	Localização genética (cM)*
DXS8378	9 330 226	20,21
DXS9902	15 233 537	32,32
DXS7132	64 572 061	90,75
DXS9898	87 683 075	101,29
DXS6809	94 824 809	108,12
DXS6789	95 336 070	108,47
DXS7133	108 928 199	118,18
GATA172D05	113 061 249	124,36
GATA31E08	140 061 921	160,54
DXS7423	149 461 561	184,19

Nota: bp - pares de bases; *- distância em cM (centiMorgans), a partir do Xp-tel usando a base NC_Chrom X.

Com o objectivo de relacionar os valores de D' e a localização cromossómica de cada microsatélite, foi seleccionado um valor representativo de D' para cada par de marcadores, correspondendo ao valor relativo ao par envolvendo os alelos mais frequentes para cada marcador. Depois, representou-se graficamente a variação de LD ao longo do segmento de cromossoma X analisado, recorrendo ao software Gold - *Graphical Overview of Linkage Disequilibrium* (Abecassis e Cooksson, 2000). O gráfico foi produzido para a população cigana e não-cigana portuguesa.

Os níveis de diferenciação interpopulacional inferidos com base em X-STRs foram avaliados através de valores de F_{ST} entre pares de populações, incluindo a população cigana portuguesa. Relativamente a X-STRs, ainda são muito escassos os dados de caracterizações populacionais, pelo que se usou como referência um estudo recentemente publicado por Gusmão *et al.* (2009).

Também se aplicou a correcção de Bonferroni para avaliar a significância dos valores de F_{ST} , e neste caso, dado se estarem a fazer 15 comparações, considerou-se $p=0,05/15=0,003$.

3. Resultados e discussão

A amostra populacional estudada neste trabalho compreendeu no total 127 indivíduos, não aparentados, de etnia cigana. O principal objectivo foi caracterizar a comunidade cigana portuguesa de forma a obter um conhecimento mais aprofundado da sua história demográfica, com auxílio de ferramentas da Genética Populacional, nomeadamente através da análise de marcadores moleculares, localizados quer em cromossomas autossómicos, quer no cromossoma X.

3.1 Análise de STRs autossómicos

3.1.1. Diversidade genética

Relativamente aos 15 marcadores autossómicos que integram o sistema multiplex AmpF/STR Identifiler alvo deste estudo, foi possível genotipar 123 indivíduos. Os alelos detectados em cada microsatélite e respectivas frequências relativas estão apresentados na tabela 7.

Na amostra analisada não se encontrou nenhum alelo novo, pois todos os alelos tinham sido anteriormente reportados noutros estudos populacionais incidindo sobre os mesmos marcadores.

CFS1PO, TPOX, D5S818 e D19S433 são os marcadores que apresentaram níveis mais baixos de diversidade. Estes 4 STRs são os únicos da bateria estudada em que alelos individuais atingiram valores de frequência superiores a 40% (CFS1PO*12=48%; TPOX*9=43,5%; D5S818*12=42,3% e D19S433*13=44,7%). CFS1PO e TPOX também se destacaram por terem o número mais baixo de alelos diferentes por *locus*.

Quando as distribuições genotípicas observadas em cada STR foram testadas para o Equilíbrio de Hardy-Weinberg, apenas foi detectado desvio estatisticamente significativo no *locus* D21S11 ($p=0,017$), assumindo o convencional limite de significância de $p=0,05$.

Mas após aplicada a correcção de Bonferroni para múltiplos testes também perdem significância os desvios registados em D21S11 relativamente às proporções genotípicas esperadas segundo o formalismo de Hardy-Weinberg.

Apesar da ausência de distorções genotípicas estatisticamente significativas, é interessante constatar que em todos os 15 STRs autossómicos estudados se tivessem registado níveis de heterozigotia esperada (H_{ESP}) sempre mais elevados do que os valores observados (H_{OBS}). Este resultado aponta para uma tendência sistemática para algum excesso de genótipos homozigóticos entre os ciganos portugueses, sendo um aspecto de alguma forma revelador da natureza endogâmica, que caracteriza em geral todas as populações ciganas.

As diversidades genéticas para cada um dos 15 STRs autossómicos na população portuguesa cigana foram comparadas com os valores previamente reportados por Amorim *et al.* (2006) para a população portuguesa não-cigana (tabela 8). Relativamente aos valores registados na população cigana portuguesa, estes variaram entre 65% e 87%, respectivamente nos *loci* TPOX e FGA, sendo concordante com o que se verificou na população não-cigana portuguesa, em que o marcador FGA foi também o marcador mais polimórfico e o TPOX o menos diverso. Porém, para a maioria dos STRs, o nível de diversidade na amostra de ciganos foi mais reduzido que o registado entre os portugueses não-ciganos.

Tabela 7. Frequências alélicas, heterozigotia observada (H_{OBS}) e heterozigotia esperada (H_{ESP}), valores de p obtidos no teste de Equilíbrio de Hardy-Weinberg para os 15 STRs autossômicos na população cigana portuguesa ($N=123$).

Alelo	D8S1179	D21S11	D7S820	CSFIPO	D3S1358	TH01	D13S317	D16S539	VWA	TPOX	D18S51	D5S818	FGA	D2S1338	D19S433
6						0,2073									
7			0,0041			0,1098									
8			0,1951			0,1992	0,2033	0,0203		0,4350*					
9			0,1545			0,1789	0,0528	0,0976		0,0610		0,0041			
9.3						0,3008									
10	0,0203		0,1260	0,2276		0,0041	0,1098	0,0691		0,1016	0,0041	0,0854			
11	0,1220		0,2561	0,2561			0,2276	0,3821		0,3902		0,2114			0,0041
12	0,1423		0,2561	0,4797*			0,3862	0,1951		0,0122	0,0772	0,4228*			0,1626
13	0,3415		0,0041	0,0325			0,0081	0,2236			0,0691	0,2602			0,4472*
13.2															0,0041
14	0,2154		0,0041	0,0041	0,0366		0,0122	0,0041	0,1748		0,2520	0,0163			0,1504
14.2															0,0203
15	0,1382				0,3171			0,0081	0,0732		0,2358				0,1220
15.2															0,0569
16	0,0203				0,2927				0,1870		0,1016			0,0285	0,0122
16.2															0,0122
17					0,1911				0,3699		0,1423			0,2195	0,0041
17.2															0,0041
18					0,1626				0,0813		0,0488			0,1341	
19									0,0772		0,0285		0,0610	0,0407	
20									0,0325		0,0203		0,2195	0,1382	
21									0,0041		0,0122		0,0772	0,0244	
21.2													0,0081		
22											0,0081		0,1707	0,0285	
22.2													0,0122		
23													0,1098	0,1057	
23.2													0,0854		
24													0,1626	0,1504	
24.2													0,0081		
25													0,0244	0,1301	
25.2		0,0041													
26													0,0569		
28		0,0732													
29		0,3211											0,0041		
30		0,1504													
30.2		0,0488													
31		0,0244													
31.2		0,0610													
32		0,0081													
32.2		0,2480													
33.2		0,0610													
H_{OBS}	0,7317	0,7642	0,6667	0,6260	0,7398	0,7398	0,7236	0,7236	0,6829	0,6016	0,7968	0,6585	0,8293	0,8537	0,6667
H_{ESP}	0,7851	0,8005	0,7942	0,6541	0,7526	0,7860	0,7461	0,7542	0,7824	0,6501	0,8403	0,7063	0,8666	0,8654	0,7351
p -HW	0,5432	0,0170	0,0587	0,0987	0,3663	0,9200	0,3381	0,2557	0,1867	0,2183	0,0572	0,2985	0,4424	0,2889	0,2397

Nota: a **negrito** encontram-se os valores de $p < 0,05$; *assinala os valores mais elevados de frequências alélicas.

Na tabela 8 também estão apresentados valores de diversidade *locus a locus* previamente publicados para outras 5 populações ciganas europeias e respectivas populações não-ciganas, habitualmente designadas por populações hospedeiras (*host populations*). Como se indica nessa tabela, foi possível compilar dados relativamente a amostras populacionais da Eslováquia (Soták *et al.*, 2008), Grécia (Deligiannidis *et al.*, 2006; Sánchez-Diz *et al.*, 2008), Hungria (Egyed *et al.*, 2000), Macedónia (Havaš *et al.*, 2007) e Croácia (Novokmet e Pavčec, 2007; Projić *et al.*, 2007).

Tabela 8. Valores de diversidade genética *locus a locus* para a população cigana portuguesa, e outras populações ciganas europeias e respectivas populações não-ciganas.

	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	VWA	TPOX	D18S51	D5S818	FGA	D2S1338	D19S433
CigPor	0,7844	0,8058	0,7908	0,6547	0,7544	0,7866	0,7461	0,7546	0,7835	0,6477	0,8384	0,7064	0,8693	0,8652	0,7237
Por	0,8114	0,8434	0,8113	0,7247	0,7873	0,7968	0,7833	0,7759	0,8105	0,6483	0,8765	0,7029	0,8675	0,8631	0,7947
CigEsl	0,8192	0,8406	0,7815	0,7123	0,7971	0,7943	0,8166	0,7689	0,7939	0,6360	0,8646	0,6724	0,8636	<i>sd</i>	<i>sd</i>
Esl	0,8143	0,8645	0,8170	0,7272	0,8133	0,7743	0,8059	0,7475	0,8222	0,5879	0,8793	0,7337	0,8740	<i>sd</i>	<i>sd</i>
CigGre	0,8336	<i>sd</i>	<i>sd</i>	<i>sd</i>	0,7511	0,8056	0,7673	<i>sd</i>	<i>sd</i>	0,6119	<i>sd</i>	0,7287	<i>sd</i>	<i>sd</i>	<i>sd</i>
Gre	0,823	0,844	0,802	0,728	0,786	0,801	0,767	0,793	0,809	0,632	0,877	0,741	0,861	0,874	0,812
CigHun	0,8258	0,8355	0,6849	<i>sd</i>	0,7599	<i>sd</i>	0,7743	<i>sd</i>	<i>sd</i>	<i>sd</i>	0,8400	0,6665	0,8554	<i>sd</i>	<i>sd</i>
Hun	0,8085	0,8552	0,7990	<i>sd</i>	0,7941	<i>sd</i>	0,7879	<i>sd</i>	<i>sd</i>	<i>sd</i>	0,8888	0,7256	0,8619	<i>sd</i>	<i>sd</i>
CigMac	0,8220	0,8559	0,8199	0,7103	0,7698	0,7946	0,7856	0,8179	0,7693	0,7164	0,8542	0,7057	0,8841	0,8851	0,7881
Mac	0,8309	0,8475	0,8081	0,7592	0,7754	0,8006	0,7844	0,7812	0,8160	0,6628	0,8860	0,7188	0,8655	0,8866	0,7895
CigCro	0,8242	0,8218	0,5465	0,6566	0,7400	0,7786	0,7962	0,8040	0,7702	0,6802	0,8081	0,6574	0,8386	0,8732	0,7945
Cro	0,7880	0,8455	0,8061	0,7224	0,7923	0,7903	0,7853	0,7727	0,8274	0,6194	0,8753	0,7100	0,8708	0,8754	0,8149

Nota: a *itálico* e a **negrito** encontram-se os valores correspondentes aos marcadores mais e menos polimórficos, respectivamente; *sd* - sem dados
CigPor: ciganos de Portugal; **Por:** população de Portugal; **CigEsl:** ciganos da Eslováquia; **Esl:** população da Eslováquia; **CigGre:** ciganos da Grécia; **Gre:** população da Grécia; **CigHun:** ciganos da Hungria; **Hun:** população da Hungria; **CigMac:** ciganos da Macedónia; **Mac:** população da Macedónia; **CigCro:** ciganos da Croácia; **Cro:** população da Croácia.

No entanto, só duas das amostras populacionais de ciganos não portugueses estavam caracterizadas para os 15 *loci* analisados neste trabalho (ciganos da Macedónia e da Croácia), já que para as restantes três apenas havia dados relativos a subconjuntos dos 15 STRs. Por essa razão, quando se calcularam valores médios de diversidade, o que implicou considerar a informação combinada entre vários *loci*, esses valores foram determinados em triplicado considerando três conjuntos de marcadores: no *set 1* só se consideraram os 4 STRs comuns a todas as amostras, os quais foram D8S1179, D3S1358,

D13S317 e D5S818; no *set 2* foram integrados os 13 marcadores do sistema CODIS, nomeadamente D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, vWA, TPOX, D18S51, D5S818 e FGA, e finalmente no *set 3* incluíram-se os 15 STRs. Os resultados apresentam-se na tabela 9.

Tabela 9. Valores médios da diversidade genética para cada população cigana e respectiva população não-cigana, considerando três conjuntos distintos de STRs autossómicos: *set 1*: restringe-se a 4 STRs (D8S1179, D3S1358, D13S317 e D5S818); *set 2*: engloba os 13 marcadores considerados para o sistema CODIS e *set 3*: inclui os 15 marcadores disponíveis no *kit* AmpF/STR Identifier.

		Ciganos	não-ciganos	% de decréscimo
Portugal	<i>set 1</i>	0,7478	0,7712	3,03
	<i>set 2</i>	0,7633	0,7877	3,10
	<i>set 3</i>	0,7674	0,7932	3,24
Eslováquia	<i>set 1</i>	0,7763	0,7918	1,95
	<i>set 2</i>	0,7816	0,7893	0,98
	<i>set 3</i>	-	-	-
Grécia	<i>set 1</i>	0,7702	0,7793	1,17
	<i>set 2</i>	-	0,7896	-
	<i>set 3</i>	-	0,7967	-
Hungria	<i>set 1</i>	0,7566	0,7790	2,88
	<i>set 2</i>	-	-	-
	<i>set 3</i>	-	-	-
Macedónia	<i>set 1</i>	0,7708	0,7774	0,85
	<i>set 2</i>	0,7927	0,7951	0,30
	<i>set 3</i>	0,7986	0,8008	0,28
Croácia	<i>set 1</i>	0,7545	0,7689	1,88
	<i>set 2</i>	0,7479	0,7850	4,73
	<i>set 3</i>	0,7593	0,7931	4,25

Como se depreende, só os valores obtidos com o *set 1* permitiram efectuar comparações envolvendo todos os grupos populacionais indicados da tabela 9. A diversidade média encontrada para a população cigana portuguesa (0,7478) foi da mesma ordem de grandeza que a estimada para os outros grupos Roma, mas representa o valor mais baixo registado entre todos os grupos ciganos, o que significa que os ciganos portugueses são os que se caracterizam por menor heterogeneidade dentro do grupo, ou por outras palavras, os que são mais idênticos entre si. Quando se compararam as diversidades médias entre populações ciganas e respectivas populações não-ciganas, verificou-se uma

diminuição sistemática dos níveis de diversidade entre os grupos ciganos, pois todas as populações hospedeiras foram mais diversas que os respectivos grupos ciganos vizinhos. Calculou-se deste modo a percentagem de decréscimo de diversidade em cada grupo cigano relativamente à população hospedeira. Os valores obtidos também se apresentam na tabela 9, e revelam que é no par ciganos/não-ciganos portugueses que se observa a maior disparidade de valores. A perda de diversidade nos ciganos portugueses é de 3%, enquanto que é no par ciganos/não-ciganos da Macedónia onde se verificam os níveis mais semelhantes entre grupos Roma e respectiva população hospedeira, a perda de diversidade nos ciganos da Macedónia é de apenas 0,85%.

Tomando em atenção a percentagem de decréscimo de heterozigotia nos pares cigano *versus* não-cigano em apreço, parece observar-se um declínio de acordo com um certo padrão geográfico (figura 8): a redução é menor em grupos de ciganos da zona dos Balcãs, nomeadamente entre os da Grécia (1,17%), Macedónia (0,85%) ou Croácia (1,88%), acentuando-se entre os ciganos estabelecidos mais a norte, como os da Hungria (2,88%) ou da Eslováquia (1,95%), apresentando o máximo decréscimo nos ciganos portugueses (3%), que habitam o extremo mais ocidental da Europa.

Este gradiente decrescente de diversidade nos Roma é coerente com o que se conhece sobre a história dos grupos Roma na Europa. De facto, o gradiente parece acompanhar a rota de migração e dispersão dos Roma através do continente (figura 8). A região dos Balcãs foi aquela em se estabeleceram os primeiros grupos que chegaram à Europa. A sua presença na região está bem documentada desde o século XIV (Fraser 1998), em que os Balcãs constituíram o foco primário da difusão dos Roma em direcção a outras regiões da Europa. Cedo ocorreram fragmentações originando sub-grupos de ciganos, que prosseguiram a sua diáspora, continuando a segmentar-se e a dispersar-se por todo o continente. Dada a localização geográfica, a Península Ibérica foi a região onde os Roma chegaram mais tarde, mas provavelmente ainda em meados do século XV.

Assim, o padrão de redução de diversidade entre os Roma poderá reflectir sucessivos efeitos de deriva genética, não sendo de estranhar observar-se uma perda de heterozigotia mais acentuada nos ciganos da Península Ibérica, pois como foram aqueles que, em princípio, resultaram de maior número de cisões, foram também os que mais acumularam efeitos de deriva.

A mesma tendência quanto à redução de diversidade em grupos Roma foi observada quando se estendeu a análise a mais marcadores, nomeadamente aos referidos *set 2* e *set 3*, mas em ambos os casos o número de grupos comparados era inferior ao disponível para o *set 1* (4 comparações para o *set 2* e apenas 3 para o *set 3*).

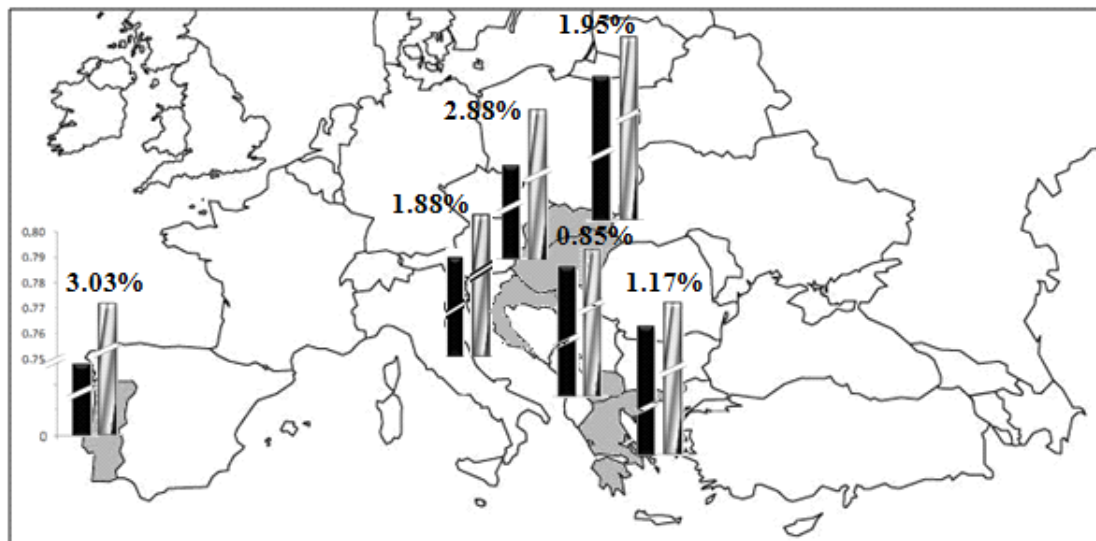


Figura 8. Padrão de decréscimo de diversidade genética entre as populações ciganas e correspondentes populações hospedeiras, com a indicação da sua localização geográfica, representado para o *set 1* (4 STRs).

Em relação a este padrão geral de decréscimo de diversidade genética, foge à sequência de encadeamento observada na população cigana da Croácia para os *sets 2* e *3*, uma vez que a respectiva taxa de redução comparativamente aos croatas não-ciganos é consideravelmente elevada, se confrontada com o que se regista nos pares ciganos/não-ciganos de localização geográfica próxima. Na amostra de ciganos croatas a redução de diversidade inferida com os *set 2* e *3*, chega a valores entre 4-5%, que são até bastante superiores ao que se observa nos ciganos portugueses. No entanto, este resultado pode ser devido a algumas peculiaridades da amostra de ciganos da Croácia estudada por Novokmet e Pavčec (2007). Em primeiro lugar, a amostra abrange 100 indivíduos todos pertencentes a uma pequena localidade situada a nordeste do país, mais concretamente à cidade de Čakovec. Parece ter sido uma amostragem muito dirigida a uma comunidade em particular e portanto pode questionar-se até que ponto terá sido possível obter amostras de indivíduos

não aparentados. Em comunidades pequenas e relativamente fechadas os níveis de consanguinidade “ambiente” tendem a ser elevados, o que por si constitui um factor de redução de diversidade intra-populacional. Em segundo lugar, segundo estes autores referem, a população que amostraram era representativa de um grupo de ciganos, minoritário na Croácia, que inclui os Roma que migraram para o nordeste do país somente no século XIX. Este movimento integrou-se numa segunda migração em massa dos Roma para algumas partes da Europa e que teve início entre finais do século XIX e inícios do século XX, com a abolição da escravatura na Roménia (Achim, 2004). Sendo assim, há um desfasamento temporal de cerca de 500 anos entre o estabelecimento na Croácia do grupo de ciganos estudado por Novokmet e Pavčec e outros Roma croatas representantes daqueles que chegaram à Península dos Balcãs e aí se estabeleceram entre os séculos XIII e XIV (Fraser, 2008). É provável que a acentuada redução de diversidade registada no único grupo de ciganos croatas caracterizado para STRs autossómicos seja também parcialmente devida ao acumular de efeitos de deriva ocorridos aquando da saída da Roménia.

Naturalmente que seria essencial dispôr de elementos não só sobre grupos de ciganos estabelecidos na Croácia há mais tempo, como sobre outros Roma e respectivas populações hospedeiras localizados na rota de dispersão pela Europa, de forma a avaliar a consistência do padrão de decréscimo de diversidade indiciada pelos dados actualmente existentes.

Na realidade o paralelismo entre a localização geográfica e a taxa de decréscimo da diversidade genética nas populações ciganas poderá ser ilusório, resultando apenas de uma qualquer distorção amostral.

A reduzida diversidade genética apresentada pelas populações ciganas tem sido constantemente verificada em diversos estudos, utilizando diferentes tipos de marcadores moleculares, reflectindo a actuação de efeitos de deriva induzidos pelos baixos efectivos populacionais, pela marcada endogamia e pela longa história de fragmentação e migração que antecedeu a actual sedentarização (Gusmão *et al.*, 2008).

3.1.2. Comparação com diferentes populações ciganas e não-ciganas

No sentido de compreender quais as afinidades genéticas entre a população cigana portuguesa e as outras populações ciganas ou não-ciganas europeias atrás referidas, foi avaliado o grau de diferenciação genética entre pares de populações, através de valores de distância genética F_{ST} , globais e *locus a locus*.

Esta medida de distância genética apenas entra em linha de conta com o número e frequência dos alelos para um dado *locus*. Em estudos que envolvem microssatélites, um tipo de marcadores com elevadas taxas de mutação, pode recorrer-se a medidas que considerem as características moleculares dos alelos. Uma medida usada habitualmente é o R_{ST} , que pondera a soma dos quadrados das diferenças do tamanho dos alelos. Neste estudo optou-se pelo F_{ST} , como já foi mencionado, por haver indicações de ser mais adequada quando estão em causa populações onde os efeitos de deriva constituíram os factores preponderantes na diferenciação entre populações ou quando fluxo génico interpopulacional é elevado.

Na tabela 10 apresentam-se as distâncias genéticas *locus a locus*, e respectivo valor de p , entre a população cigana portuguesa e as outras cinco populações ciganas europeias. Os valores de F_{ST} entre a população portuguesa cigana e não-cigana variam entre 0,011% e 3,92%, e são estatisticamente significativos em 8 dos 15 microssatélites, mesmo após aplicação da correcção de Bonferroni para múltiplos testes. São resultados que evidenciam uma diferenciação significativa entre as duas populações.

Tabela 10. Valores de distância genética (F_{ST}) *locus a locus* entre a população cigana portuguesa e as cinco populações ciganas europeias, e respectivos valores de significância estatística (p), para um desvio padrão $\leq 0,01$.

<i>Locus</i>	<i>CigEsl</i>		<i>CigGre</i>		<i>CigHun</i>		<i>CigMac</i>		<i>CigCro</i>	
	F_{ST}	p	F_{ST}	p	F_{ST}	p	F_{ST}	p	F_{ST}	p
D8S1179	-0,0021	0,5791	0,0295	0,0009	0,0028	0,1875	0,0084	0,0898	0,0080	0,0977
D21S11	0,0020	0,2647	-	-	-0,0027	0,7256	0,0104	0,0518	0,0039	0,2129
D7S820	0,0323	0,0009	-	-	0,0607	0,0000	0,0109	0,0576	0,1395	0,0000
CSF1PO	0,0003	0,3584	-	-	-	-	0,0223	0,0225	0,0063	0,1690
D3S1358	0,0011	0,3262	-0,0028	0,56345	0,0143	0,0205	-0,0078	0,9697	0,0011	0,3447
TH01	0,0014	0,3203	-0,0016	0,5215	-	-	0,0017	0,2910	-0,0079	0,9697
D13S317	0,0093	0,0429	-0,0006	0,4287	0,0026	0,2412	-0,0024	0,5615	0,0335	0,0039
D16S539	0,0194	0,0127	-	-	-	-	0,0196	0,0127	0,0352	0,0001
VWA	0,0010	0,3359	-	-	-	-	-0,0009	0,4590	0,0450	0,0000
TPOX	-0,0028	0,5576	-0,0062	0,7393	-	-	0,0284	0,0176	-0,0055	0,7285
D18S51	0,0092	0,0342	-	-	0,0216	0,0009	0,0131	0,0244	0,0263	0,0020
D5S818	0,0045	0,1973	-0,0014	0,4512	0,0106	0,0566	-0,0036	0,6045	0,0235	0,0215
FGA	0,0090	0,0391	-	-	0,0173	0,0019	0,0083	0,0557	0,0208	0,0039
D2S1338	-	-	-	-	-	-	0,0063	0,0811	0,0204	0,0020
D19S433	-	-	-	-	-	-	0,0031	0,2158	0,0137	0,0449

Nota a negrito encontram-se os valores de p significativos, após correcção de Bonferroni.
CigEsl: ciganos da Eslováquia; **CigGre:** ciganos da Grécia; **CigHun:** ciganos da Hungria; **CigMac:** ciganos da Macedónia; **CigCro:** ciganos da Croácia.

Quando os ciganos portugueses são comparados com outros grupos de ciganos, a diferenciação mais acentuada que se encontra é relativamente aos ciganos da Croácia. Em geral os valores de F_{ST} *locus a locus* entre ciganos de Portugal e os da Croácia estão na gama dos mais altos registados, e em 5 dos 15 *loci* analisados as distâncias são estatisticamente significativas. A diferenciação com os ciganos da Hungria também é clara. Embora para esta comparação apenas houvesse dados para 8 STRs, em 3, os F_{ST} s foram estatisticamente significativos. As maiores afinidades dos ciganos portugueses parecem ocorrer com os ciganos da Macedónia, visto que nenhuma das distâncias inferidas *locus a locus* atinge níveis de significância estatística.

Também foram calculadas as distâncias genéticas entre a população cigana portuguesa e as 6 populações europeias não-ciganas (tabela 11).

Tabela 11. Valores de distância genética (F_{ST}) *locus a locus* entre a população cigana portuguesa e as cinco populações não-ciganas europeias, e respectivos valores de significância estatística (p), para um desvio padrão $\leq 0,001$.

<i>Locus</i>	<i>Por</i>		<i>Gre</i>		<i>Hun</i>		<i>Mac</i>		<i>Cro</i>		<i>Esl</i>	
	F_{ST}	p	F_{ST}	p	F_{ST}	p	F_{ST}	p	F_{ST}	p	F_{ST}	p
D8S1179	0,0011	0,2822	0,0033	0,1563	0,0031	0,1924	0,0013	0,3164	-0,0011	0,4756	-0,0033	0,7061
D21S11	0,0258	0,0000	0,0137	0,0010	0,0326	0,0000	0,0046	0,1602	0,0328	0,0000	0,0177	0,0049
D7S820	0,0154	0,0000	0,0189	0,00293	0,0231	0,0010	0,0185	0,0166	0,0180	0,0049	0,0138	0,0234
CSF1PO	0,0203	0,0010	0,0246	0,0010	-	-	0,0191	0,0371	0,0075	0,0967	0,0176	0,0273
D3S1358	0,0013	0,2705	-0,0004	0,4434	0,0014	0,2940	-0,0053	0,8018	0,0022	0,2402	0,0127	0,0293
TH01	0,0032	0,1397	0,0023	0,2188	-	-	-0,0026	0,5781	-0,0002	0,4336	0,0043	0,1699
D13S317	0,0241	0,0000	0,0117	0,02148	0,0215	0,0010	0,0126	0,0576	0,0195	0,0088	0,0197	0,0068
D16S539	0,0093	0,0127	0,0100	0,02637	-	-	0,0051	0,1572	0,0136	0,0117	0,0111	0,0606
VWA	0,0179	0,0000	0,0075	0,04883	-	-	0,0194	0,0088	0,0329	0,0000	0,0116	0,0293
TPOX	0,0111	0,0283	0,0266	0,0020	-	-	0,0056	0,1709	0,0179	0,0186	0,0389	0,0029
D18S51	0,0148	0,0000	0,0176	0,0000	0,0119	0,0039	0,0142	0,0098	0,0104	0,0137	0,0069	0,0762
D5S818	0,0172	0,0059	0,0054	0,1270	0,0108	0,0371	-0,0019	0,4785	0,0090	0,0869	0,0193	0,00880
FGA	0,0130	0,0020	0,0235	0,0000	0,0102	0,0176	0,0100	0,0449	0,0131	0,0039	0,0128	0,0137
D2S1338	0,0037	0,07129	0,0037	0,1104	-	-	0,0010	0,3565	-0,0005	0,4717	-	-
D19S433	0,0393	0,0000	0,0411	0,0000	-	-	0,0426	0,0010	0,0543	0,0000	-	-

Nota: a **negrito** encontram-se os valores de p significativos após correcção de Bonferroni.

Por: população de Portugal; **Gre:** população da Grécia; **Hun:** população da Hungria **Mac:** população da Macedónia; **Cro:** população da Croácia; **Esl:** população da Eslováquia.

Relativamente às comparações dos ciganos portugueses com populações não-ciganas, o maior número de valores individuais de F_{ST} por *locus* estatisticamente significativos foi registado com Portugal (8 em 15 *loci*), a Grécia (6 em 15 *loci*), Hungria (3 em 8 *loci*) e a Croácia (3 em 15 *loci*). Não se registaram distâncias significativas com a Eslováquia, e com a Macedónia, apenas num *locus* se observou um valor significativo (tabela 11).

Fazendo uma análise global destes resultados, não parecem existir diferenças substanciais entre o grau de diferenciação dos ciganos portugueses relativamente a outros grupos Roma assim como também ao nível da diferenciação comparativamente a populações não-ciganas. Por outro lado, também o nível de diferenciação genética dos ciganos portugueses em relação aos portugueses não-ciganos é da mesma ordem do que os distancia de outras populações europeias não-ciganas.

Se esta análise *locus a locus* forneceu pistas sobre o relacionamento genético entre as populações comparadas, tem o inconveniente de envolver uma quantidade de

informação pouco abrangente. Por este motivo foram determinados os valores globais de F_{ST} entre pares populacionais (tabela 12).

No sentido de excluir factores de enviesamento dos resultados obtidos, foram consideradas nesta análise apenas as populações ciganas e respectivas populações *host* com dados referentes a 13 STRs. Por outro lado, de forma a integrar a diversidade autossómica dos ciganos num contexto mais amplo, também se considerou uma amostra da população da Turquia (Çakir *et al.*, 2003), outra do Afeganistão (Hohoff *et al.*, 2006) e quatro amostras populacionais da Índia. Quanto a estas quatro, referem-se a 4 grandes regiões geográficas da Índia assinaladas no mapa da figura 9: *Andhra Pradesh* (Noroeste da Índia), *Maharashtra* (Índia Oriental), *Karnataka* (Sudoeste da Índia) e *Chhattisgarh* (Índia Central) e cada uma compreende resultados compilados em vários trabalhos (Bindu *et al.*, 2007; Gaikwad e Kashyap, 2003; Rajkumar e Kashyap, 2003; Sarkar e Kashyap, 2002).

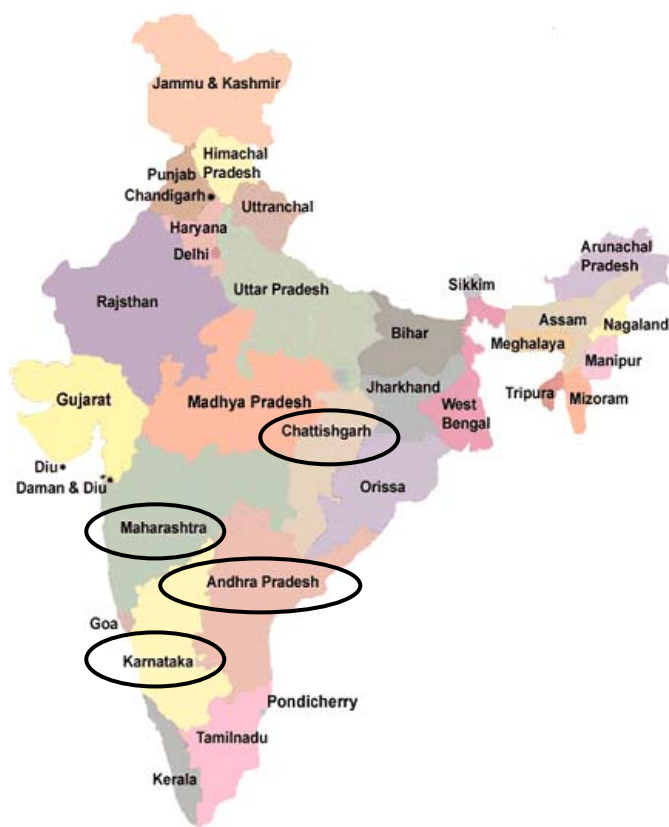


Figura 9. Mapa da Índia com a localização geográfica (marcadas com círculo preto) das populações usadas para comparação com a população cigana portuguesa em estudo.

O critério utilizado na selecção das populações não europeias fundamentou-se na história da origem e diáspora dos Roma. Tudo indica que a Índia terá sido o local de origem das primeiras populações ciganas que, antes de chegarem à Europa, tiveram que percorrer um extenso corredor migratório. A Turquia e o Afeganistão localizam-se na rota seguida pelos Roma após terem abandonado a Índia (Liégeois, 1989).

Na tabela 12 encontram-se reunidos os valores de F_{ST} globais para os pares de populações envolvidos na análise.

Tabela 12. Valores de F_{ST} (abaixo da diagonal) entre as populações comparadas para os 13 STRs (excepto D2S1338 e D19S433), e respectivos valores de significância estatística (p) (acima da diagonal), para um desvio padrão $\leq 0,001$.

	CigPor	Por	CigEsl	Esl	CigMac	Mac	CigCro	Cro	Turq	Afeg	Andh*	Mah*	Karn*	Chhat*
CigPor	*	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Por	0,0148	*	0,0000	0,0284	0,0000	0,1115	0,0000	0,0065	0,0013	0,0000	0,0000	0,0000	0,0000	0,0000
CigEsl	0,0100	0,0094	*	0,0000	0,0001	0,0032	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Esl	0,0173	0,0024	0,0091	*	0,0000	0,0621	0,0000	0,0539	0,0069	0,0000	0,0000	0,0000	0,0000	0,0000
CigMac	0,0127	0,0121	0,0089	0,0175	*	0,0249	0,0000	0,0001	0,0018	0,0002	0,0003	0,0000	0,0000	0,0002
Mac	0,0118	0,0016	0,0057	0,0035	0,0063	*	0,0000	0,0824	0,1349	0,0189	0,0001	0,0000	0,0003	0,0002
CigCro	0,0300	0,0346	0,0196	0,0342	0,0317	0,0307	*	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Cro	0,0167	0,0028	0,0119	0,0028	0,0156	0,0029	0,0360	*	0,0039	0,0000	0,0000	0,0000	0,0000	0,0000
Turq	0,0197	0,0057	0,0108	0,0072	0,0117	0,0024	0,0341	0,0068	*	0,0391	0,0006	0,0000	0,0033	0,0003
Afeg	0,0150	0,0062	0,0102	0,0087	0,0099	0,0045	0,0312	0,0077	0,0031	*	0,0000	0,0000	0,0003	0,0000
Andh*	0,0183	0,0151	0,0153	0,0199	0,0120	0,0113	0,0392	0,0179	0,0085	0,0099	*	0,0002	0,0929	0,0044
Mah*	0,0264	0,0217	0,0242	0,0250	0,0212	0,0210	0,0433	0,0246	0,0147	0,0161	0,0059	*	0,0030	0,0006
Karn*	0,0165	0,0147	0,0141	0,0189	0,0132	0,0115	0,0350	0,0172	0,0080	0,0076	0,0015	0,0054	*	0,0215
Chhat*	0,0146	0,0163	0,0138	0,0203	0,0116	0,0124	0,0336	0,0183	0,0103	0,0103	0,0040	0,0064	0,0035	*

Nota: * populações da Índia.

CigPor: ciganos de Portugal; **Por:** população de Portugal; **CigEsl:** ciganos da Eslováquia; **Esl:** população da Eslováquia; **CigGre:** ciganos da Grécia; **Gre:** população da Grécia; **CigHun:** ciganos da Hungria; **Hun:** população da Hungria; **CigMac:** ciganos da Macedónia; **Mac:** população da Macedónia; **CigCro:** ciganos da Croácia; **Cro:** população da Croácia; **Turq:** população da Turquia; **Afeg:** população do Afeganistão; **Andh:** população de Andhra Pradesh; **Mah:** população de Maharashtra; **Karn:** população de Karnataka; **Chhat:** população de Chhattisgarh.

Analisando os valores de F_{ST} , verificou-se que a distância média entre pares de populações ciganas foi de 0,019 enquanto que a distância média entre grupo cigano e respectiva população hospedeira foi de 0,017. Tratam-se de valores da mesma ordem de grandeza, o que significa que o grau de diferenciação entre grupos Roma é semelhante ao que distancia grupo Roma de não-Roma do mesmo país. Como termo de comparação, registre-se que a distância média entre populações europeias não-ciganas é de 0,003, o que

traduz uma diferenciação cerca de 6 vezes inferior à que se observa quando está envolvida uma população cigana.

Na sua grande maioria, os valores de F_{ST} apresentados na tabela 12 são estatisticamente significativos, mesmo aplicando o factor de correcção de Bonferroni ($p=0,003$), sendo sempre significativos os valores encontrados entre a população cigana portuguesa e as restantes populações.

Foi entre o par ciganos portugueses-croatas que se detectou um distanciamento maior (3,0%), e as menores diferenciações registaram-se entre os ciganos portugueses relativamente aos da Eslováquia e Macedónia (1,0% e 1,3%, respectivamente). Estes resultados mostram grande coerência com os valores de F_{ST} marcador a marcador apresentados anteriormente (tabela 10).

Para obter uma representação visual das afinidades entre as várias populações comparadas, foi realizada uma análise *Multidimensional Scalling* (MDS) a duas dimensões com base nos valores de distâncias genéticas apresentados na tabela 12. A análise foi efectuada com o programa Statistica versão 8.0, e resultou no *plot* apresentado na figura 10.

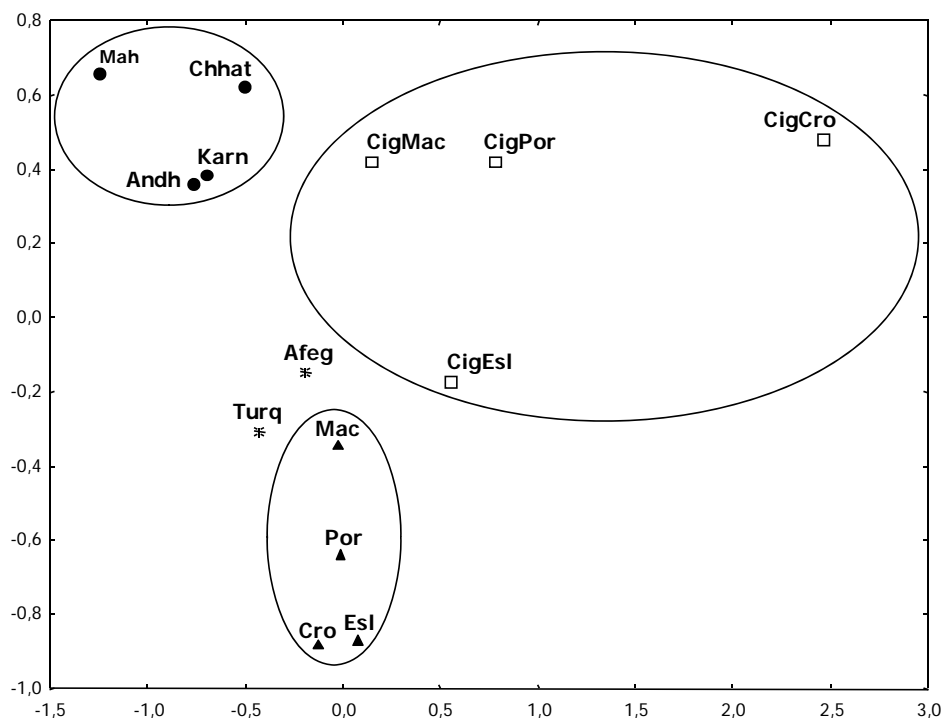


Figura 10. Plot resultante da Análise Multidimensional (MDS) baseada em F_{ST} s entre diversas populações Roma e não-Roma.

CigPor: ciganos de Portugal; **Por:** população de Portugal; **CigEsl:** ciganos da Eslováquia; **Esl:** população da Eslováquia; **CigGre:** ciganos da Grécia; **Gre:** população da Grécia; **CigHun:** ciganos da Hungria; **Hun:** população da Hungria; **CigMac:** ciganos da Macedónia; **Mac:** população da Macedónia; **CigCro:** ciganos da Croácia; **Cro:** população da Croácia; **Turq:** população da Turquia; **Aveg:** população do Afeganistão; **Andh:** população de Andhra Pradesh; **Mah:** população de Maharashtra; **Karn:** população de Karnataka; **Chhat:** população de Chhattisgarh.

Na figura 10 destacam-se claramente três grandes grupos de populações: a) um que engloba as populações da Índia, b) um segundo que reúne as populações europeias ciganas e c) um terceiro onde se agrupam as populações europeias não-ciganas. A população do Afeganistão ocupa uma posição intermédia em relação ao *cluster* da Índia e da Europa não-cigana, e quanto à Turquia, são nítidas as afinidades com as populações europeias não-ciganas, para os marcadores analisados.

Focando a atenção no *cluster* das populações ciganas, sobressai a dispersão e distanciamento espacial dos grupos Roma, sobretudo quando comparado com o que se observa no *cluster* das populações europeias não-ciganas, muito mais aglomerado, apesar de compartilhar com o anterior o mesmo enquadramento geográfico. A homogeneidade genética das populações europeias não-ciganas é aliás, um aspecto bem documentado.

Ainda em relação ao agrupamento dos Roma, é de salientar a localização periférica dos ciganos da Croácia, que são os que evidenciam maior nível de diferenciação genética no conjunto das populações ciganas. O isolamento relativo da população croata cigana, bem visível no *plot* da figura 10, sugere que efeitos de deriva muito intensos devem ter modelado o perfil genético deste grupo, acabando por originar uma maior diferenciação relativamente a outras populações ciganas. Esta interpretação é sustentada pelas características deste grupo de ciganos (Projić *et al.*, 2007), já referidas para justificar a marcada diminuição de variabilidade genética que registam comparativamente à população croata não-cigana.

Uma forma alternativa de representar relações entre grupos populacionais é através de árvores filogenéticas, opção a que também se recorreu neste trabalho. Os valores de F_{ST} globais, apresentadas na tabela 12, foram utilizados para construir uma árvore pelo método de *Neighbor Joining*, incluído no programa Phylip versão 3.65, e a respectiva visualização foi posteriormente efectuada com o programa FigTree versão 1.2.2. A árvore obtida apresenta-se na figura 11.

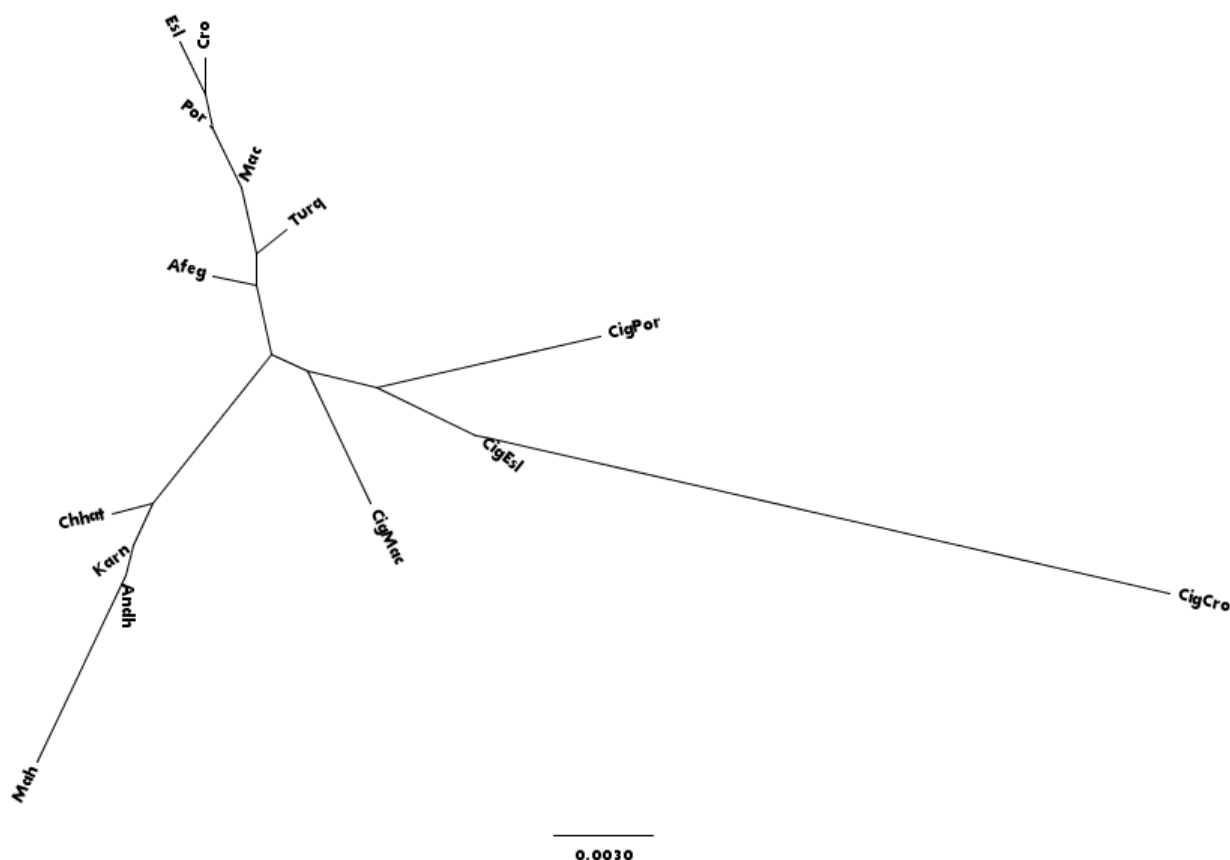


Figura 11. Árvore obtida por *Neighbour Joining* a partir da matriz de distâncias genéticas da tabela 12. O tamanho dos ramos é proporcional à distância genética entre populações, e a escala da figura é dada em unidades de F_{ST} .

CigPor: ciganos de Portugal; **Por:** população de Portugal; **CigEsl:** ciganos da Eslováquia; **Esl:** população da Eslováquia; **CigGre:** ciganos da Grécia; **Gre:** população da Grécia; **CigHun:** ciganos da Hungria; **Hun:** população da Hungria; **CigMac:** ciganos da Macedónia; **Mac:** população da Macedónia; **CigCro:** ciganos da Croácia; **Cro:** população da Croácia; **Turq:** população da Turquia; **Afeg:** população do Afeganistão; **Andh:** população de Andhra Pradesh; **Mah:** população de Maharashtra; **Karn:** população de Karnataka; **Chhat:** população de Chhattisgarh.

Na representação em árvore as relações interpopulacionais são basicamente coincidentes com as observadas na análise MDS: distinguem-se bem os mesmos três grandes grupos de populações, nomeadamente da Índia, da Europa e de ciganos.

Sobressai ainda a grande diferenciação entre populações ciganas, como indica o longo tamanho dos ramos relativos aos grupos Roma. Destaca-se em particular a extensão do ramo referente à população cigana croata, característica que denota claros efeitos de deriva. O posicionamento da Turquia e do Afeganistão relativamente às outras populações é também muito semelhante ao que se observa no MDS da figura 10.

Um terceiro e último método foi usado para avaliar as relações entre as populações em causa, desta vez independente de uma medida prévia de distância genética. Assim, com base apenas nas frequências alélicas para cada um dos 13 STRs autossómicos, foi realizada uma Análise de Componentes Principais (PCA), recorrendo igualmente a uma rotina do *software* Statistica versão 8.0. Os 2 primeiros componentes produzidos na análise projectam-se na figura 12.

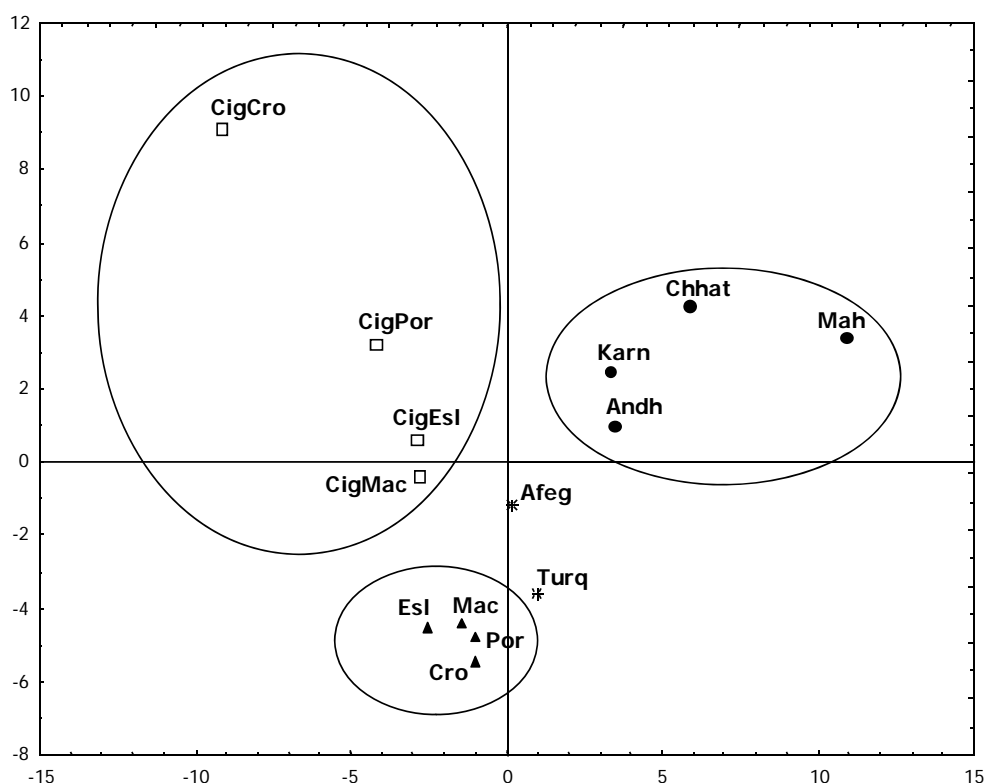


Figura 12. Plot resultante da Análise de Componentes Principais (PCA) baseada em frequências alélicas de 13 STRs autossómicos, incluindo diversas populações Roma e não-Roma.

CigPor: ciganos de Portugal; **Por:** população de Portugal; **CigEsl:** ciganos da Eslováquia; **Esl:** população da Eslováquia; **CigGre:** ciganos da Grécia; **Gre:** população da Grécia; **CigHun:** ciganos da Hungria; **Hun:** população da Hungria; **CigMac:** ciganos da Macedónia; **Mac:** população da Macedónia; **CigCro:** ciganos da Croácia; **Cro:** população da Croácia; **Turq:** população da Turquia; **Afeg:** população do Afeganistão; **Andh:** população de Andhra Pradesh; **Mah:** população de Maharashtra; **Karn:** população de Karnataka; **Chhat:** população de Chhattisgarh.

Os resultados obtidos com PCA apontam para um cenário global de relacionamento interpopulacional totalmente concordante com o deduzido pelas duas abordagens anteriores (figuras 10 e 11). Entre outros aspectos, os três tipos de análise revelam a grande heterogeneidade existente entre os grupos Roma, que denuncia os fortes efeitos de deriva a

que estas populações estiveram sujeitas. Apesar disso, todos as populações ciganas tendem a agrupar-se entre si, formando um *cluster* coerente e distinto de outros grupos de populações. É um padrão que revela um elevado grau de partilha de património genético entre grupos Roma, e que reforça a ideia de que os ciganos da Europa tiveram uma origem comum.

3.1.3. Estimativas de miscigenação nas populações ciganas

Existem várias evidências que indicam que o sub-continente indiano foi o local de origem das primeiras vagas de migração dos proto-Roma. Naturalmente que ao longo de muitos séculos de fragmentação e diáspora, o *pool* genético dos grupos Roma foi-se alterando relativamente ao ancestral, sob o efeito de diversos factores, entre os quais, como já se referiu, fortes efeitos de deriva genética resultantes do reduzido efectivo populacional dos Roma. Paralelamente, quer durante a dispersão quer após a sedentarização, o contacto com outras populações proporcionou muitas oportunidades para trocas genéticas, e o influxo de genes contribuiu certamente para a remodelação do perfil genético dos grupos Roma.

Entendeu-se, por isso, tentar avaliar o grau de mistura no *pool* genético das populações actuais dos Roma. Para o efeito utilizou-se o programa Admix 2.0 que permite obter estimativas da contribuição de n populações parentais para uma dada população misturada, recorrendo a um modelo de miscigenação descrito por Bertorelle e Excoffier (1998). Na realidade, a complexa história demográfica dos ciganos não se encaixa em nenhum modelo simples de mistura populacional, como é o de Bertorelle e Excoffier, mas devido à indisponibilidade de modelações de mistura mais complexas, avançou-se com a análise na expectativa de obter aproximações grosseiras à intrincada história de mistura populacional registada entre os Roma.

Como se compreende, a escolha das populações parentais foi em parte condicionada pela disponibilidade de dados populacionais para os 13 marcadores autossómicos em que se baseou a análise. Face a essa limitação e atendendo à história de migração e dispersão dos Roma, foram considerados 3 grupos parentais: 1) Europa (resultante do *pooling* das amostras europeias não-ciganas, dado que se integram num

mesmo *cluster*, como se pode ver nas figuras 10, 11 e 12); 2) Médio-Oriente (representado pela junção das amostras da Turquia e Afeganistão, as únicas populações do noroeste asiático, um importante “corredor migratório” dos Roma, com dados publicados); e 3) Índia (aglomerando as quatro amostras das regiões indianas consideradas nas análises anteriormente descritas). As frequências génicas das “mega-populações” parentais foram obtidas através da média ponderada pelo efectivo populacional das frequências compiladas da literatura para populações individuais contidas em cada grupo.

Quando se estimam coeficientes de miscigenação baseados em STRs, o programa Admix 2.0 permite levar ou não em conta as diferenças moleculares entre alelos. Neste caso, optou-se por não se atender às características moleculares mas apenas às diferenças de frequências alélicas. Visto terem sido usados marcadores com elevadas taxas de mutação, como são os STRs, este procedimento tem sido considerado o mais apropriado para obter estimativas de mistura quando estão em jogo populações em que efeitos de deriva genética foram preponderantes na sua história evolutiva (Dupanloup *et al.*, 2004).

Finalmente, no modelo de mistura implementado no Admix 2.0 também é possível entrar em consideração com a antiguidade do acontecimento de miscigenação. No caso das populações ciganas, seria irrealista fixar uma única idade para o processo de mistura, dado que este decorreu ao longo de séculos durante os quais também se alteraram as populações que forneceram os principais aportes genéticos. Por essa razão, foram efectuados três ensaios independentes, assumindo as idades prováveis de algumas etapas do processo de mistura populacional de que resultaram possivelmente os Roma actuais, nomeadamente, 700, 1000 e 1700 anos. A idade de 1700 anos, foi a presumida para o início do êxodo populacional da Índia, de acordo com algumas fontes que situam as primeiras migrações ainda no século III (Fraser, 1998). Essa data representa, assim, aquela a partir da qual os proto-Roma começaram a incorporar características genéticas de origem não-indiana. Relativamente às outras duas alternativas de idade consideradas, procurou-se contemplar o limite inferior (1000 anos) e o superior (700 anos) dos contactos com populações europeias. O primeiro, 1000 anos, fundamenta-se em alguns testemunhos que dão conta da chegada dos Roma ao Império Bizantino, tão cedo quanto o século XI (Fraser 1998). O segundo, 700 anos, baseia-se no conhecimento de que a passagem para os Balcãs foi mais tardia, havendo fortes indícios de que a fixação na região remonta ao século XIV.

Os valores obtidos para as proporções de mistura nas populações ciganas de Portugal, Eslováquia, Macedónia e Croácia encontram-se indicados na tabela 13.

Tabela 13. Estimativas de miscigenação e respectivo valor padrão (σ) para as populações ciganas europeias considerando três idades diferentes (700, 1000 e 1700 anos).

População	700 anos			1000 anos			1700 anos		
	Europa	Turq + Afeg	Índia	Europa	Turq + Afeg	Índia	Europa	Turq + Afeg	Índia
CigPor	0,394	0,436	0,170	0,318	0,474	0,208	0,314	0,420	0,266
σ	0,043	0,068	0,032	0,019	0,028	0,015	0,009	0,011	0,007
CigEsl	0,366	0,403	0,230	0,308	0,443	0,249	0,310	0,403	0,287
σ	0,039	0,061	0,029	0,018	0,025	0,015	0,008	0,010	0,007
CigMac	0,486	0,246	0,268	0,380	0,355	0,264	0,347	0,361	0,292
σ	0,040	0,063	0,031	0,019	0,028	0,015	0,009	0,012	0,007
CigCro	0,343	0,476	0,182	0,304	0,474	0,223	0,311	0,415	0,274
σ	0,044	0,068	0,034	0,020	0,029	0,017	0,009	0,012	0,008

Numa primeira observação global dos resultados, sobressai a consistência dos valores encontrados nos diferentes Roma, pois em todos são muito semelhantes as estimativas das contribuições relativas dos três grupos parentais, qualquer que seja o cenário temporal modelado.

Os resultados desta análise sugerem que foi substancial a contribuição europeia para o perfil genético dos ciganos actuais, já que as respectivas estimativas rondam sempre percentagens superiores a 30%. Talvez as melhores aproximações quanto a esse contributo sejam as inferidas com as idades mais recentes, ou seja de 1000 e 700 anos, pois balizam o início dos contactos entre populações ciganas e europeias, durante o século XI, e o estabelecimento de ligações mais regulares com o enraizamento de populações ciganas em território europeu, a partir do século XIV. Assim, o componente genético de origem europeia nos ciganos portugueses é da ordem dos 39% ou 32%, assumindo, respectivamente, 700 ou 1000 anos para o início da mistura. O grupo Roma que apresenta um componente europeu mais elevado, é o da Macedónia – 49% ou 36% consoante se assumam os 700 ou os 1000 anos, e os Roma onde a influência europeia parecer ter sido

menor são os da Croácia, mas ainda assim é de 34% ou 30%, conforme se considere os mesmos 700 ou 1000 anos, a fracção de origem europeia do seu *pool* genético.

As pequenas diferenças entre grupos Roma, quanto às contribuições genéticas resultantes de possíveis influxos europeus, foram razoavelmente capturadas pela análise PCA previamente descrita e que resultou no *plot* apresentado na figura 12. Nessa figura pode observar-se que os Roma da Macedónia são os que mais se aproximam do *cluster* que engloba as populações europeias não-ciganas, enquanto que os Roma da Croácia são os que mais se distanciam do mesmo *cluster*.

Em relação à contribuição ancestral da Índia, as estimativas obtidas representam, em todos os grupos de ciganos e todas as modelações temporais, os valores mais baixos das possíveis contribuições parentais. Esses valores tendem a decrescer com a diminuição da idade assumida para o acontecimento de mistura, mas pela mesma ordem de ideias seguida para o componente europeu, quanto ao componente indiano, a idade mais antiga, 1700 anos, parecer ser a que facultará a melhor estimativa deste componente. Essa idade remete para a escala temporal em que ocorreu a saída dos primeiros proto-ciganos da Índia, representando, portanto, aquela a partir da qual a ancestralidade original começou a ser diluída. Neste pressuposto, o componente indiano que ainda permanece retido no *pool* genético dos actuais Roma, não ultrapassa os 27% nos ciganos de Portugal e da Croácia, e 29% nos da Macedónia e da Eslováquia. A escassez de vestígios indianos pode significar que, ao longo dos anos, grande parte da composição ancestral dos Roma, de cariz original indiano, foi sendo substituída por outros aportes populacionais. Mas atendendo a que a formação dos Roma é um processo relativamente recente, provavelmente desenrolado nos últimos 1700 anos, uma taxa de renovação genética tão acentuada pode também indicar que foi muito pequeno o número de indianos envolvido na vaga migratória inicial que saiu do sub-contidente, e que esteve na origem dos Ciganos actuais.

Por último, refiram-se as contribuições atribuídas à terceira população parental presumida, o Médio-Oriente, e que agrupa, como já se descreveu, amostras da Turquia e do Afeganistão. Na maioria dos grupos ciganos, a proporção de mistura com esta população ronda valores da ordem dos 40%. Exceptuam-se os Roma da Macedónia, em que a proporção correspondente é um pouco mais baixa, caindo para 25% a 36%. De qualquer forma, todas as proporções são relativamente elevadas, indicando que contribuições

importantes para o *pool* autossómico dos Roma podem ter resultado de contactos com populações afins à aqui referida como Médio-Oriente.

Parece, pois, que uma fracção substancial do genoma autossómico dos ciganos foi adquirida durante a travessia do noroeste asiático, na rota de migração entre a Índia e a Europa. Foi um longo corredor não só de passagem mas também de estabelecimento de comunidades, o que poderá explicar que a influência de populações da região ficasse profundamente marcada no património genético das populações ciganas.

3.2. Análise de X-STRs

3.2.1. Diversidade genética

A amostra da população cigana portuguesa foi também analisada para os 10 microssatélites do cromossoma X, de acordo com o processo já mencionado em Material e Métodos. As frequências alélicas e valores de diversidade genética (H_{ESP}) para cada marcador estão apresentados na tabela 14.

Quanto a estes marcadores, não se testou o equilíbrio de Hardy-Weinberg uma vez que a amostra analisada só incluía indivíduos do sexo masculino, sendo, portanto, todos hemizigóticos para os marcadores do X.

Todos os alelos encontrados em cada um dos 10 X-STR já tinham sido reportados noutros estudos populacionais baseados na mesma bateria de marcadores (eg. Gusmão *et al.*, 2008). Este decaplex de X-STRs, foi originalmente optimizado tendo em vista aplicações forenses, pelo que um dos critérios de selecção dos marcadores tinha sido a elevada diversidade genética.

Não surpreende, portanto, que na amostra de ciganos portugueses se tivessem detectado níveis de polimorfismo bastante elevados para qualquer um dos 10 X-STRs. Como consta da tabela 14, os valores de diversidade génica por *locus* variaram entre um máximo de 0,8319, registado no marcador DXS6809, e um mínimo de 0,5998, no marcador DXS7133. Neste último *locus*, a reduzida diversidade deve-se em parte provavelmente ao facto de um dos seus alelos, o alelo 9, atingir a frequência de 0,5854,

representando o valor mais elevado entre as frequências alélicas individuais dos STRs analisados.

Tabela 14. Frequências alélicas e valores de diversidade génica (H_{ESP}) em 10 X-STRs na população cigana portuguesa ($N=123$).

Alelo	DXS8378	DXS9898	DXS7133	GATA31E08	GATA172D05	DXS7423	DXS6809	DXS7132	DXS9902	DXS6789
6					0,1951					
7				0,0650						
8				0,0081	0,1463					
8.3		0,2276								
9	0,0732		0,5854*	0,4634	0,0325			0,0081		
10	0,2683		0,1138	0,1626	0,2602				0,3252	
11	0,3333	0,0813	0,2114	0,2764	0,3577			0,0081	0,3984	
11.1									0,0163	
12	0,2846	0,3496	0,0650	0,0244	0,0081			0,1951	0,2602	
13	0,0325	0,2195	0,0244			0,0650		0,2520		
14	0,0081	0,1220				0,2195		0,3008		
15						0,5041		0,1707		0,0650
16						0,2114		0,0650		0,0081
17										0,0325
19										0,0081
20										0,3984
21										0,2846
22										0,1382
23										0,0244
24										0,0407
29							0,0569			
30							0,0244			
31							0,2439			
32							0,1951			
33							0,2276			
34							0,1057			
35							0,0894			
36							0,0081			
37							0,0488			
H_{ESP}	0,7354	0,7625	0,5998	0,6831	0,7498	0,6541	0,8319	0,7808	0,6731	0,7396

Nota: * assinala o valor mais elevado de frequência alélica.

Na tabela 15 apresentam-se valores de diversidade genética *locus a locus* e valores de diversidade média para o decaplex de X-STRs, na amostra estudada e noutras populações. Nestas não aparece mais nenhuma população cigana porque, tanto quanto foi possível apurar, não se encontraram dados publicados para X-STRs que proporcionassem um quadro comparativo de interesse. Mesmo para populações não-ciganas são ainda escassos dados relativos a X-STRs, e por isso optou-se por utilizar nas comparações a compilação de dados populacionais relativos ao decaplex de X-STRs recentemente publicada por Gusmão *et al.* (2009), que inclui apenas populações ibéricas e latino-americanas. Para além destas, considerou-se ainda, a título comparativo, uma amostra dos Karimojong do Uganda, por ser uma das poucas populações caracterizadas para o conjunto completo de 10 X-STRs (Gomes *et al.*, 2009b).

Começando por comparar os níveis médios de diversidade para o X-decaplex entre ciganos e não-ciganos portugueses, verifica-se uma redução de heterozigotia nos primeiros relativamente aos segundos (0,7210 *versus* 0,7446). Assim, com um conjunto diferente de STRs reforça-se o que se observou com os STRs autossómicos, ou seja, uma perda de diversidade na população cigana comparativamente a outros portugueses.

A explicação adiantada para o registado nos STRs autossómicos aplica-se igualmente aos marcadores do X, e reside na combinação de factores demográficos como o baixo efectivo populacional dos ciganos, a manutenção das tradicionais práticas endogâmicas e num passado marcado por sucessivas fragmentações populacionais. São factores que propiciam a ocorrência de fortes efeitos de deriva genética e que tendem a induzir diminuição da diversidade nas populações, tal como se verifica nos ciganos portugueses.

Para o X-decaplex, a taxa de redução de diversidade nos ciganos foi de 3,17% comparativamente aos outros portugueses, sendo um valor muito semelhante ao observado para os STRs autossómicos que constava de 3,03%, 3,10% e 3,24%, considerando 4, 13 e 15 STRs, respectivamente, conforme se apresentou na tabela 9.

Tabela 15. Valores de diversidade genética *locus a locus* e valores médios para a população cigana portuguesa e outras populações com dados publicados.

	DXS8378	DXS9898	DXS7133	GATA31E08	GATA172D05	DXS7423	DXS6809	DXS7132	DXS9902	DXS6789	10 X-STRs
CigPort	0,7354	0,7625	0,5998	0,6831	0,7498	0,6541	0,8319	0,7808	0,6731	0,7396	0,7210
Port	0,6957	0,7718	0,6731	0,7936	0,8013	0,6928	0,8078	0,7432	0,7115	0,7555	0,7446
Aço	0,7029	0,7789	0,6621	0,8012	0,8119	0,7009	0,8105	0,7396	0,7202	0,7182	0,7446
Gal	0,6984	0,7799	0,6744	0,7946	0,8004	0,6893	0,8122	0,7495	0,6986	0,7224	0,7420
Cant	0,7029	0,7775	0,6497	0,7997	0,7983	0,7181	0,8014	0,7595	0,6894	0,6992	0,7396
C Rica	0,6813	0,7783	0,6106	0,7906	0,8173	0,7406	0,7974	0,7484	0,6845	0,7041	0,7353
Ant	0,6364	0,7894	0,6048	0,7843	0,7957	0,7055	0,7787	0,7572	0,6746	0,7320	0,7259
Par	0,6857	0,7833	0,6725	0,7991	0,8012	0,7087	0,8210	0,7635	0,7215	0,7622	0,7519
R Jan	0,7030	0,7900	0,6910	0,8221	0,8261	0,6797	0,8323	0,7916	0,7274	0,8295	0,7693
S Paul	0,6871	0,7867	0,7052	0,8120	0,8063	0,6903	0,8453	0,7685	0,7112	0,8093	0,7622
M Gros	0,6911	0,7893	0,6922	0,8007	0,8250	0,7191	0,8160	0,7725	0,7112	0,7950	0,7612
E Ríos	0,6876	0,7654	0,5935	0,7793	0,7445	0,7561	0,7851	0,7433	0,7027	0,7840	0,7341
Mis	0,6951	0,7240	0,7389	0,8028	0,7795	0,6420	0,8215	0,8042	0,6401	0,7361	0,7384
Córd	0,6937	0,7564	0,6055	0,8028	0,7678	0,7439	0,8201	0,7510	0,6721	0,7593	0,7373
B Aires	0,6936	0,7637	0,6328	0,7973	0,7925	0,6796	0,7999	0,7792	0,7039	0,7239	0,7366
R Neg	0,7080	0,7566	0,6287	0,7536	0,7899	0,7285	0,8393	0,7924	0,6876	0,7404	0,7425
Karim	0,6834	0,7508	0,5134	0,8036	0,7885	0,6880	0,8449	0,7565	0,7346	0,8538	0,7417

Nota: a **negrito** estão assinalados os valores médios de diversidade genética dos 10 X-STRs; **CigPor**: ciganos de Portugal; **Port**: população portuguesa; **Aço**: população dos Açores; **Gal**: população da Galiza; **Cant**: população da Cantábria; **C Rica**: população da Costa Rica; **Ant**: população de Antioquia; **Par**: população do Paraná; **R Jan**: população do Rio de Janeiro; **S Paul**: população de São Paulo; **M Gros**: população de Mato Grosso; **E Ríos**: população de Entre Ríos; **Mis**: população de Misiones; **Córd**: população de Córdoba; **B Aires**: população de Buenos Aires; **R Neg**: população de Rio Negro; **Karim**: população de Karimojong.

Teoricamente, e uma vez que o número efectivo de cromossomas X é cerca de $\frac{3}{4}$ do dos autossomas, seria de esperar uma redução de diversidade mais acentuada para marcadores do X do que em autossómicos, admitindo, por um lado, cruzamentos ao acaso entre igual número efectivo de homens (N_h) e de mulheres (N_m) e, por outro, idênticas taxas de mutação nos marcadores. Dado o menor número de cromossomas X, os efeitos de deriva são mais acentuados em marcadores do cromossoma X que em autossómicos, sendo de prever que a razão entre a taxa de redução de diversidade no X comparativamente à dos autossomas se aproximasse de ~75%.

Quanto às taxas de mutação, de momento ainda não existem dados sobre os X-STRs que permitam comparações com o que já se encontra publicado para os autossómicos. Mas não é de admitir que qualquer possível diferença influenciasse muito os níveis de diversidade do X comparativamente a autossomas, pois a história dos ciganos não parece ser suficientemente antiga para que os efeitos da mutação se reflectissem nos padrões de diversidade.

Assim, o facto de se terem verificado níveis idênticos de redução de heterozigotia em STRs do cromossoma X e de autossomas, correspondendo a uma razão de taxas de 0,98 (3,17/3,24) considerando para autossomas o valor inferido com o maior número de marcadores), pode ser uma indicação de que a deriva actuou numa população em que o número efectivo de homens e mulheres era diferente. Mais concretamente, como a razão de taxas é muito superior à razão prevista de 0,75, ou seja, como os efeitos de deriva na diversidade ligada ao cromossoma X foi menos forte do que esperado assumindo $N_h=N_m$, há indicações de que o número efectivo de mulheres ciganas tem sido superior ao número efectivo de homens.

Naturalmente que esta interpretação tem que ser encarada com bastante reserva. Primeiro por se apoiar num número restrito de marcadores, o que compromete o valor do resultado. Segundo, porque se baseia numa observação registada num único grupo de ciganos. Seria importante confirmar noutros Roma se a tendência também se verifica, pois não há motivos para pensar que as histórias demográficas de homens e mulheres variem muito entre grupos de ciganos. Finalmente, porque se considerou um modelo simples de deriva, em que dificilmente se encaixa a história complexa dos ciganos. Os resultados obtidos com STR autossómicos, apresentados no ponto anterior, são claros na indicação de que a mistura populacional foi um processo fundamental ao longo da história dos ciganos. Pode pois questionar-se até que ponto os influxos génicos terão contribuído para neutralizar a redução de diversidade por efeitos de deriva. Certamente que esses influxos concorreram para restaurar pelo menos parte da diversidade perdida por deriva, mas a taxa de recuperação tanto se pode ter exercido ao nível da diversidade do cromossoma X, como da dos autossomas, não parecendo, por isso, ser justificação para o que se observou.

Na realidade o resultado observado pode até ter sido ocasional e não reflectir uma tendência real. No entanto, também não se pode excluir desde já a hipótese de que o número efectivo de mulheres na população cigana ter sido de facto superior ao dos

homens. É difícil especular sobre as causas que poderão ter induzido nos ciganos uma redução relativa do efectivo populacional de homens, cuja explicação pode residir em diversos factores. Uma maior variância no sucesso reprodutivo dos homens comparativamente ao das mulheres implicaria uma diminuição do número efectivo de indivíduos do sexo masculino. Infelizmente não se dispõem de elementos que permitam fundamentar esta possibilidade, mas o assunto poderia representar um excelente pretexto para, futuramente, desenvolver em colaboração com especialistas doutras áreas, nomeadamente da Sociologia ou da História, investigações que pudessem trazer algumas achegas à questão.

Regressando à análise da tabela 15 e olhando agora para o conjunto inteiro de populações, verifica-se que os níveis médios de diversidade para o X-decaplex variam relativamente pouco de população para população, pois a gama de variação encontra-se apenas entre ~73% (na população de Antioquia) e ~77% (na população do Rio de Janeiro). A diversidade intrapopulacional nas amostras de Portugal Continental, Açores, Galiza e Cantábria, é muito semelhante, rondando os 74%. Os níveis de diversidade mais elevados encontram-se em populações brasileiras como as do Rio de Janeiro (0,7693), São Paulo (0,7622) e do Mato Grosso (0,7612), ou seja, em populações em que foi grande a influência de mistura populacional recente (Gusmão *et al.*, 2009).

É sintomático que entre o número considerável de populações listadas na tabela 15, sejam os ciganos portugueses que apresentam nível de diversidade genética mais baixa (0,7210).

A maioria das populações listadas, pouca relevância tem no contexto da história dos ciganos. Mas como já foi explicado, não houve grande margem para escolha, face à escassez de dados para o conjunto de marcadores do cromossoma X estudados. Seguramente que quando houver para este decaplex dados publicados para outras populações ciganas, ou populações envolvidas na sua formação, será possível extrair mais informação através de análises comparativas.

3.2.2. Padrão de desequilíbrio de ligação (LD)

Dado que os 10 marcadores em causa se encontram localizados no mesmo cromossoma, e uma vez que o sistema decaplex foi desenvolvido tendo em vista essencialmente aplicações forenses, tem sido uma preocupação analisar em diferentes populações os padrões de LD, pois a observação de associações significativas entre marcadores implicaria alguma redundância de informação, com consequente redução de informatividade do sistema. Por outras palavras, a presença de associações invalidaria a presunção de independência entre *loci*, uma presunção especialmente crítica quando estão em causa aplicações do tipo mencionado.

No sentido de analisar o padrão de LD na amostra de ciganos portugueses, foram realizados testes exactos de associação entre todos os pares possíveis de marcadores incluídos no X-decaplex, ou seja, entre 45 pares de marcadores, e os correspondentes valores de p bem como respectivos valores de desvio padrão apresentam-se na tabela 16. Foram ainda avaliados os níveis de LD entre todos os pares possíveis de alelos de diferentes *loci*. A medida utilizada foi o parâmetro D' , pois tratando-se do determinante gamético (D) normalizado pelo valor máximo de D para um dado jogo de frequências alélicas, permite comparar níveis de LD entre pares de alelos com frequências distintas. Na tabela 16 apresentam-se os valores de D' , mas apenas referentes à combinação dos alelos mais comuns para cada par de *loci*.

Para proporcionar um contexto comparativo, foram efectuadas idênticas análises numa amostra de portugueses não-ciganos, e os resultados obtidos também constam da tabela 16.

Tabela 16. Valores de p de testes exactos de desequilíbrio de ligação entre pares de *loci*, respectivo desvio padrão (σ) e parâmetro D' , para os alelos mais frequentes de cada par de marcadores, nas populações de portugueses ciganos e não-ciganos. A distância genética entre marcadores apresenta-se em centiMorgans (cM).

Pares de <i>loci</i>	<i>Ciganos</i>			<i>não-ciganos</i>			Distância
	valor de p	σ	D'	valor de p	σ	D'	
DXS8378-DXS9898	0,1469	0,0022	-0,0930	0,7415	0,0028	-0,0714	81,08
DXS8378-DX7133	0,5658	0,0031	-0,1250	0,0266	0,0012	-0,0187	97,97
DXS8378-GATA31E08	0,6726	0,0033	0,0455	0,9220	0,9220	0,1464	140,33
DXS8378-GATA172D05	0,4829	0,0040	0,0127	0,0616	0,0014	0,0865	104,15
DXS8378-DXS7423	0,3035	0,0033	-0,2258	0,5055	0,0032	-0,1808	163,98
DXS8378-DXS6809	0,3500	0,0021	-0,1000	0,5398	0,0023	0,1724	87,91
DXS8378-DXS7132	0,1487	0,0016	0,1081	0,1889	0,0022	-0,0116	70,54
DXS8378-DXS9902	0,0753	0,0016	0,0270	0,0058	0,0008	0,0088	12,11
DXS8378-DXS6789	0,2019	0,0015	-0,1429	0,6148	0,0019	-0,0492	88,26
DXS9898-DXS7133	0,0563	0,0013	0,0465	0,6417	0,0017	-0,1595	16,89
DXS9898-GATA31E08	0,4787	0,0028	0,1332	0,9982	0,9982	-0,0174	59,25
DXS9898-GATA172D05	0,3133	0,0035	-0,2199	0,5410	0,0035	-0,0919	23,07
DXS9898-DXS7423	0,7312	0,0033	0,0621	0,8817	0,0014	-0,0221	82,9
DXS9898-DXS6809	0,0806	0,0017	-0,6186	0,4989	0,0027	-0,1595	6,83
DXS9898-DXS7132	0,1841	0,0026	0,0858	0,6385	0,0025	-0,0923	10,54
DXS9898-DXS9902	0,0073	0,0008	-0,1827	0,5743	0,0031	0,0296	68,97
DXS9898-DXS6789	0,0000	0,0000	-0,4746	0,1372	0,0014	-0,0738	7,18
DXS7133-GATA31E08	0,0384	0,0014	-0,0499	0,2551	0,0021	-0,0617	42,35
DXS7133-GATA172D05	0,0250	0,0007	-0,2623	0,8339	0,0021	0,0636	6,17
DXS7133-DXS7423	0,7461	0,0027	0,0275	0,3773	0,0041	0,0562	66,01
DXS7133-DXS6809	0,0032	0,0005	-0,0889	0,9787	0,0008	-0,1189	10,07
DXS7133-DXS7132	0,7437	0,0022	-0,0135	0,8993	0,0022	-0,0308	27,43
DXS7133-DXS9902	0,7284	0,0036	0,0001	0,4910	0,0021	-0,0946	85,87
DXS7133-DXS6799	0,0465	0,0010	-0,0935	0,2486	0,0015	-0,2089	9,71
GATA31E08-GATA172D05	0,4348	0,0038	0,1105	0,5364	0,0027	-0,0397	36,18
GATA31E08-DXS7423	0,6165	0,0036	-0,0603	0,2722	0,0043	-0,2295	23,65
GATA31E08-DXS6809	0,6782	0,0033	0,0682	0,1020	0,0015	-0,2494	52,42
GATA31E08-DXS7132	0,3594	0,0039	-0,1252	0,0812	0,0013	0,0614	69,79
GATA31E08-DXS9902	0,5368	0,0038	0,0111	0,6924	0,0028	-0,1379	128,22
GATA31E08-DXS6789	0,8209	0,0027	0,0111	0,9940	0,0008	0,0977	52,07
GATA172D05-DXS7423	0,0037	0,0005	0,4042	0,1120	0,0018	-0,1473	59,83
GATA172D05-DXS6809	0,3022	0,0024	-0,0682	0,6280	0,0020	-0,0507	16,24
GATA172D05-DXS7132	0,2973	0,2214	0,0742	0,1394	0,0017	0,0184	33,61
GATA172D05-DXS9902	0,1609	0,0024	0,1689	0,0085	0,0007	-0,3186	92,04
GATA172D05-DXS6789	0,2856	0,0028	-0,2013	0,1619	0,0020	0,1679	15,89
DXS7423-DXS6809	0,6479	0,0022	0,0590	0,2302	0,0020	0,0809	76,07
DXS7423-DXS7132	0,0098	0,0005	0,1280	0,9490	0,0018	-0,0299	93,44
DXS7423-DXS9902	0,3111	0,0036	-0,0283	0,5393	0,0031	0,0531	151,87
DXS7423-DXS6789	0,7765	0,0019	-0,1903	0,6487	0,0035	0,0558	75,72
DXS6809-DXS7132	0,0096	0,0007	-0,4459	0,2943	0,0012	0,0098	17,37
DXS6809-DXS9902	0,7671	0,0037	-0,1633	0,9204	0,0013	0,0946	75,8
DXS6809-DXS6789	0,0000	0,0000	0,5014	0,1568	0,0017	0,0523	0,35
DXS7132-DXS9902	0,5712	0,0034	-0,0502	0,7348	0,0022	-0,0816	58,43
DXS7132-DXS6789	0,0668	0,0013	-0,0502	0,6960	0,0017	-0,1024	17,72
DXS9902-DXS6789	0,6862	0,0038	-0,0779	0,0134	0,0005	-0,1174	76,15

Nota: a **negrito** assinalam-se os valores de p estatisticamente significativos após correcção de Bonferroni; em *itálico* indicam-se os valores de distância genética correspondentes aos pares de p estatisticamente significativos.

Na população cigana portuguesa, dois dos testes exactos de associação entre *loci* foram altamente significativos, mesmo após aplicação da correcção de Bonferroni para múltiplos testes, que neste caso se considerou $p < 0,001$ como limite de confiança estatística. Os valores significativos correspondiam aos pares DXS9898-DXS6789 e DXS6809-DXS6789, com $p \leq 0,0000$ em ambos os testes. Tendo em conta a localização no cromossoma X dos 10 STRs (figura 2 da Introdução), verifica-se que os 3 *loci* envolvidos nestas associações fazem parte do mesmo grupo de ligação (DXS9898-DXS6809-DXS6789) estando geneticamente muito próximos entre si: a distância entre DXS9898 e DXS6789 é de apenas 7,18 cM, sendo bastante menor a distância entre DXS6809 e DXS6789, que é de 0,35 cM. Relativamente a outros *loci* em estreita ligação, como DXS9898 e DXS6809 (6,83 cM), ou DXS7133 e GATA172D05 (6,17 cM), os valores de p dos testes de associação são baixos mas não chegam a atingir significância estatística (ver tabela 16).

Já na população portuguesa não-cigana, nenhuma associação significativa foi detectada entre os 45 pares de marcadores envolvendo os mesmos 10 STRs, e assim, pelo menos para o cromossoma X, não há sinais de LD significativo nesta população.

Quanto ao par DXS6809-DXS6789, este trabalho sobre os ciganos portugueses não é o primeiro em que se encontrou associação significativa entre os 2 STRs. Anteriormente, Szibor *et al.* (2005), reportam associação entre os mesmos marcadores numa amostra populacional da Alemanha, e Gomes *et al.* (2007) também detectam associação entre DXS6809 e DXS6789 numa amostra de Afro-Americanos dos EUA, ainda que neste caso, a associação perdesse significado estatístico após aplicação da correcção de Bonferroni. Pelo facto destes marcadores estarem muito próximos entre si (0,35 cM), compreende-se a observação de sinais de LD, mais ou menos fortes, em populações com histórias muito distintas.

A ligação por si não gera LD. As causas principais de geração de LD são essencialmente de ordem demográfica, como subestruturação ou mistura populacional, baixo efectivo populacional, entre outras, ainda que a elevada mutabilidade dos STRs possa dar alguma contribuição para o aparecimento de LD.

Mas uma vez estabelecido um certo nível de LD, o facto de haver ligação estreita entre marcadores, retarda substancialmente o seu decaimento ao longo do tempo, pois a baixa frequência de recombinação pouco contribui para quebrar as associações entre alelos

de diferentes *loci*. Assim, os sinais de LD podem perdurar muito tempo nas populações. Estando em causa o cromossoma X e comparativamente a autossomas, espera-se até que esses sinais persistam mais tempo uma vez que a taxa média de recombinação no cromossoma X é cerca de $\frac{2}{3}$ da dos autossomas.

O que se encontrou nos ciganos portugueses que não se tinha observado em qualquer outra população foi ainda no grupo de ligação DXS9898-DXS6809-DXS6789, a existência de associação não só entre DXS6809 e DXS6789, mas também entre DXS9898 e DXS6789. Isto significa que nos ciganos portugueses o LD se estende por uma porção mais extensa do cromossoma X, comparativamente ao que se tinha verificado noutras populações com sinais positivos de LD.

A primeira ilação destes resultados, é que em estudos envolvendo simultaneamente os 3 X-STRs, DXS9898, DXS6809 e DXS6789, não se deve considerar independente a informação alélica marcador a marcador, sendo conveniente atender à sua combinação haplotípica. Por essa razão, e para facilitar futuros estudos comparativos, apresentam-se na tabela 17 os haplótipos detectados na população de ciganos portugueses, bem como as respectivas frequências, cujas estimativas se obtiveram por contagem directa, uma vez que só se analisaram homens.

Tabela 17. Frequências haplotípicas e respectivos valores de desvio padrão (σ) envolvendo os microsatélites DXS9898, DXS6809 e DXS6789, na população cigana portuguesa.

Haplótipo	DXS9898	DXS6809	DXS6789	Frequência	σ
1	11	31	20	0,0407	0,0179
2	13	31	20	0,0325	0,0161
3	13	31	21	0,0163	0,0115
4	12	31	20	0,0244	0,0140
5	83	32	20	0,0081	0,0081
6	83	31	21	0,0163	0,0115
7	83	32	19	0,0081	0,0081
8	83	31	20	0,0488	0,0195
9	12	33	21	0,0325	0,0161
10	11	34	21	0,0081	0,0081
11	12	34	24	0,0244	0,0140
12	83	35	22	0,0163	0,0115
13	14	31	23	0,0163	0,0115
14	13	32	22	0,0163	0,0115
15	13	34	20	0,0163	0,0115
16	13	35	22	0,0163	0,0115
17	83	32	22	0,0081	0,0081
18	14	34	24	0,0081	0,0081
19	12	32	21	0,0650	0,0223
20	12	29	15	0,0488	0,0195
21	12	33	20	0,0244	0,0140
22	12	35	22	0,0407	0,0179
23	14	33	20	0,0244	0,0140
24	14	31	21	0,0081	0,0081
25	83	30	22	0,0081	0,0081
26	12	34	21	0,0325	0,0161
27	13	32	21	0,0325	0,0161
28	83	33	22	0,0081	0,0081
29	13	33	17	0,0244	0,0140
30	11	33	21	0,0163	0,0115
31	12	33	15	0,0081	0,0081
32	12	35	20	0,0081	0,0081
33	12	31	21	0,0081	0,0081
34	12	37	20	0,0163	0,0115
35	13	37	20	0,0081	0,0081
36	83	34	24	0,0081	0,0081
37	13	32	20	0,0081	0,0081
38	83	32	15	0,0081	0,0081
39	13	34	21	0,0081	0,0081
40	14	31	20	0,0244	0,0140
41	83	33	20	0,0407	0,0179
42	13	33	20	0,0325	0,0161
43	83	33	17	0,0081	0,0081
44	83	32	21	0,0244	0,0140
45	11	30	21	0,0081	0,0081
46	14	37	20	0,0163	0,0115
47	83	37	20	0,0081	0,0081
48	13	33	22	0,0081	0,0081
49	14	35	20	0,0081	0,0081
50	12	29	16	0,0081	0,0081
51	12	30	22	0,0081	0,0081
52	14	36	23	0,0081	0,0081
53	83	31	22	0,0081	0,0081
54	14	32	21	0,0081	0,0081
55	11	32	20	0,0081	0,0081

A observação de associações significativas entre DXS6809 e DXS6789 ou entre DXS9898 e DXS6789 na amostra de ciganos e a sua ausência na amostra de portugueses não ciganos, reflecte-se naturalmente, nos valores de D' para cada par de marcadores observados nas duas populações. Já se justificou a opção pela estatística D' para medir níveis de LD, mas convém agora referir que é um parâmetro que varia entre -1 e 1, e em que $D'=0$ significa não haver associação entre *loci*.

Ora como se pode ver na tabela 16, enquanto que nos portugueses não-ciganos D' foi de -0,0738 e 0,0523, para DXS9898-DXS6789 e DXS6809-DXS6789, respectivamente, os valores correspondentes na amostra de ciganos foram substancialmente mais elevados, chegando a -0,4746, no primeiro par de STRs, e 0,5014, no segundo.

De forma a obter uma perspectiva mais nítida sobre as diferenças globais de níveis de LD nas duas amostras populacionais comparadas, foi utilizada a estratégia descrita por Leite *et al.* (2009), que consistiu em calcular o *baseline* médio de LD populacional, obtendo-o através da média dos valores absolutos de D' entre pares de *loci* que não façam parte do mesmo grupo de ligação. Para este efeito, consideraram-se os grupos de ligação descritos por Szibor *et al.* (2003b), conforme estão assinalados no ideograma do cromossoma X da figura 2 da Introdução.

Na população cigana o *baseline* de LD foi de 0,1037, sendo um valor ligeiramente superior ao registado da população portuguesa não cigana igual a 0,0941. Calculou-se então o valor médio de LD entre pares de *loci* do grupo de ligação 2 (ver figura 2), ou seja o grupo a que pertencem os *loci* onde se detectaram associações significativas.

Enquanto que na população cigana o valor médio de LD entre STRs deste grupo foi de 0,2163, cerca do dobro do *baseline* médio de LD nos ciganos (0,1037), na população não cigana o LD médio para o mesmo grupo de ligação foi apenas de 0,0948, diferindo pouco do *baseline* médio da própria população (0,0941). Esta análise não foi efectuada para os outros grupos de ligação devido ao reduzido número de STRs que integra cada grupo.

Na figura 13 apresentam-se os padrões de LD na população portuguesa cigana e na não cigana. Na representação utilizada, os valores de D' entre todos os pares de STRs são projectados entrando em consideração com a localização e distância física entre os marcadores.

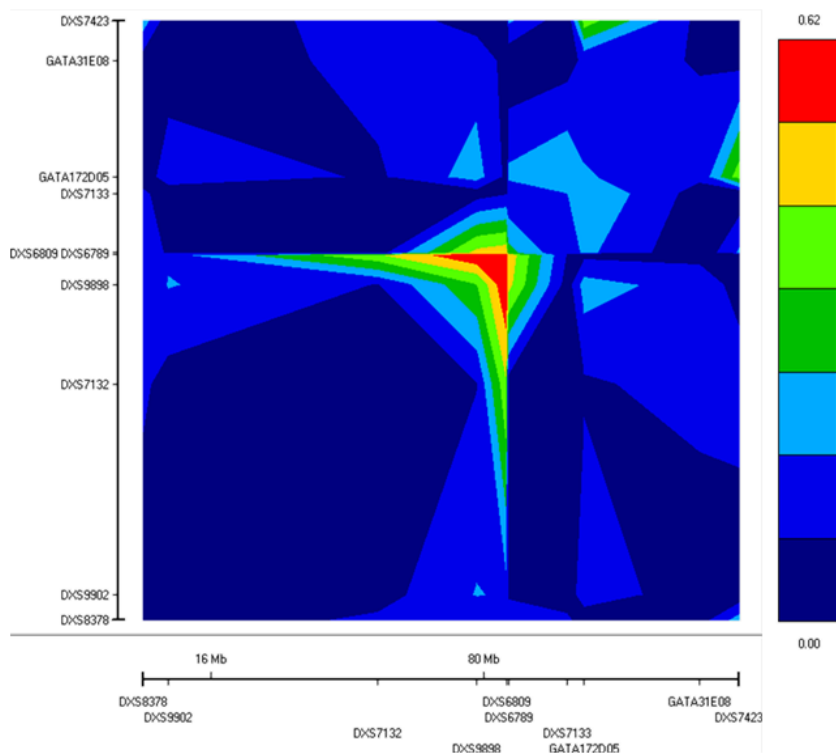
Para além de se poder observar que os níveis de LD são consideravelmente mais elevados na população cigana que na não-cigana, as projecções de D' revelam ainda que na população cigana se nota bem um decaimento de LD em função da distância física (figura 13 A), enquanto que na outra população não se observa qualquer padrão consistente de variação de LD (figura 13 B).

A mistura populacional é outra causa importante de geração de LD. E no caso dos ciganos, desde a saída da Índia, que a sua história tem sido marcada por mistura populacional. Portanto será outro factor que, combinado com a deriva genética, deverá ter contribuído para o considerável nível de LD detectado entre os ciganos portugueses.

O padrão de LD resultante de mistura é altamente dependente da dinâmica do processo de mistura. Recentemente, Pfaff *et al.* (2001), simularam a evolução populacional de LD ao longo do tempo, assumindo a) um modelo simples de mistura, *Hybrid-Isolation* (HI), em que intensa mistura ocorre apenas numa única geração, e b) um outro de *Continuous Gene Flow* (CGF), em que a mistura ocorre a uma taxa contínua mas reduzida em cada geração. No primeiro modelo, o declínio de LD inicia-se após o acontecimento de mistura e progride a uma taxa que é, naturalmente, dependente da fracção de recombinação. No modelo CGF, a quantidade de LD aumenta nas primeiras gerações à medida que influxo génico gera mais desequilíbrio, depois estabiliza durante várias gerações, e então começa a decair a uma taxa muito mais lenta que a observada no modelo HI, de tal modo que se continuam a detectar associações para *loci* em estreita ligação ao fim de várias dezenas de gerações após início da mistura. Destes dois modelos, o *Continuous Gene Flow* parece ser aquele em que melhor se encaixam as populações ciganas, e ainda assim não incorpora o facto de as populações que forneceram os influxos génicos aos ciganos terem alterado ao longo do tempo, isto é, não entra em linha de conta com as variações sofridas pelas populações que foram contribuindo para o *pool* genético das populações ciganas.

De qualquer modo, de momento afigura-se muito difícil discriminar na população cigana o LD gerado por mistura, daquele resultante de efeitos de deriva. Eventualmente, uma cobertura mais fina de marcadores do cromossoma X talvez pudesse contribuir para obter um quadro mais preciso do padrão populacional de LD. Por outro lado, seria também importante conhecer as frequências génicas dos mesmos marcadores nas possíveis populações parentais que contribuíram para o *gene pool* dos ciganos, uma vez que a

A.



B.

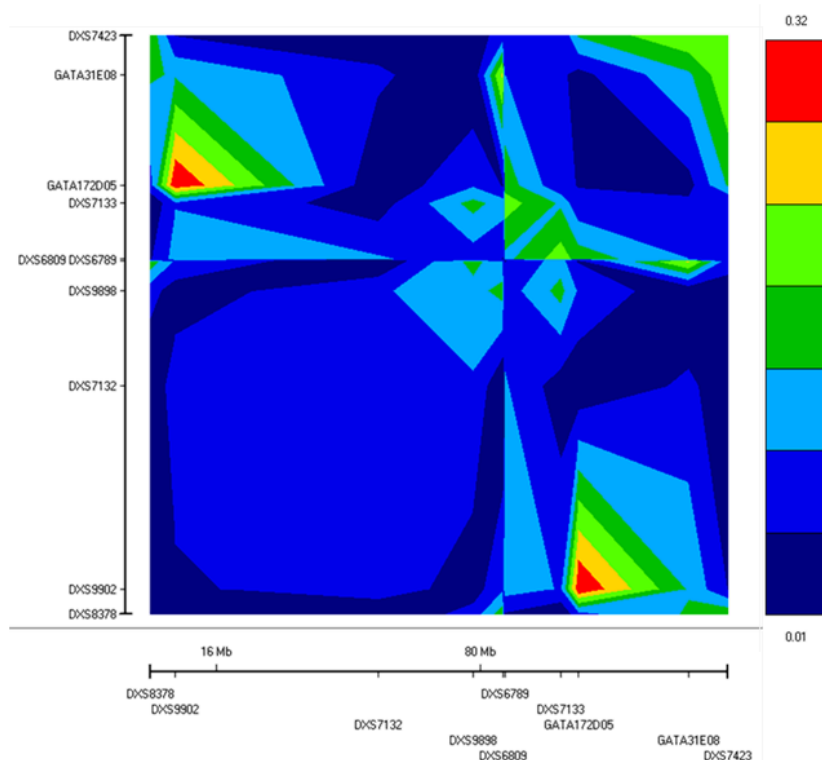


Figura 13 – Padrões de LD na população portuguesa cigana (A) e não-cigana (B) baseados nos valores de D' . Na escala dos eixos encontram-se os 10 X-STRs analisados, tendo em conta as respectivas distâncias físicas em Mb (tabela 6). O *software* Gold representa graficamente cada par de *loci* por cores: vermelho, verde e azul significando grau forte, moderado e fraco, respectivamente de LD.

extensão de LD é muito dependente da diferença de frequências entre as populações envolvidas na mistura.

3.2.3. Diversidade interpopulacional

Com o objectivo de avaliar as afinidades genéticas entre a população cigana portuguesa e outras populações com dados publicados para o X-decaplex, foram calculadas distâncias genéticas (F_{ST}) entre cada par de populações. As populações consideradas foram as constantes do já referido trabalho de Gusmão *et al.* (2009), que para além de Portugal, Galiza e Cantábria incidiu sobre 13 populações latino-americanas: Costa Rica, Antioquia (Colômbia), Paraná, Rio de Janeiro São Paulo e Mato Grosso do Sul (Brasil), Entre Ríos, Misiones, Córdoba, Buenos Aires e Rio Negro (Argentina). Também foi considerada uma amostra da população açoriana (Silva *et al.*, 2009) e outra dos Karimojong do Uganda estudada por Gomes *et al.* (2009).

Uma vez que foram detectadas associações significativas entre DXS9898, DXS6809 e DXS6789, no cálculo das distâncias genéticas não se consideraram como independentes as frequências alélicas para cada um destes STRs, usando-se antes as frequências haplotípicas apresentadas na tabela 17. Os valores de F_{ST} entre os ciganos portugueses e as referidas populações encontram-se na tabela 18.

Como se pode ver, todas as distâncias genéticas são altamente significativas, à excepção da distância relativamente à população de Entre Ríos, que perde significância após correcção de Bonferroni.

O grau de diferenciação dos ciganos relativamente aos outros portugueses é muito semelhante ao registado em relação às populações dos Açores e de Espanha (Galiza e Cantábria), o que se entende dada a ausência de diferenças significativas entre as populações portuguesas e as espanholas (Gusmão *et al.*, 2009).

Entre as populações da América Latina, aquelas das quais os ciganos portugueses se diferenciam mais (valores de F_{ST} entre 2,3 a 3%) são as populações brasileiras do Rio de Janeiro, São Paulo e Mato Grosso, possivelmente por serem as mais miscigenadas das populações latino americanas consideradas. Recorde-se, que estas três populações foram as que apresentaram os níveis de diversidade mais elevados para os X-STRs (ver tabela 15).

Tabela 18. Valores de F_{ST} entre a população cigana portuguesa e outras populações para os 10 X-STRs e respectivos valores de significância estatística (p), para um desvio padrão $\leq 0,001$.

População	F_{ST}	p
Port	0,02028	0,00000
Aço	0,02141	0,00000
Gal	0,01743	0,00000
Cant	0,01376	0,00000
C Rica	0,01925	0,00000
Ant	0,02446	0,00000
Par	0,01466	0,00000
R Jan	0,03894	0,00000
S Paul	0,02599	0,00000
M Gros	0,02356	0,00000
E Ríos	0,01021	0,02344
Mis	0,01582	0,00000
Córd	0,01041	0,00098
B Aires	0,01190	0,00000
R Neg	0,02138	0,00000
Karim	0,07662	0,00000

Nota: a **negrito** – indica valores de p estatisticamente significativos após correcção de Bonferroni.

CigPor: ciganos de Portugal; **Port:** população portuguesa; **Aço:** população dos Açores; **Gal:** população da Galiza; **Cant:** população da Cantábria; **C Rica:** população da Costa Rica; **Ant:** população de Antioquia; **Par:** população do Paraná; **R Jan:** população do Rio de Janeiro; **S Paul:** população de São Paulo; **M Gros:** população de Mato Grosso; **E Ríos:** população de Entre Ríos; **Mis:** população de Misiones; **Córd:** população de Córdoba; **B Aires:** população de Buenos Aires; **R Neg:** população de Rio Negro; **Karim:** população de Karimojong.

Por fim é de referir a grande distância genética dos ciganos relativamente aos Karimojong (7,7%), que de certa forma era previsível, dado não compartilharem uma história recente comum.

Como já foi referido, a maioria destas populações comparadas pouca relevância tem no contexto da história dos ciganos. Ainda assim, e não havendo dados populacionais mais pertinentes, entendeu-se que os disponíveis não poderiam deixar de ser mencionados neste trabalho.

Naturalmente que seria interessante expandir esta análise comparativa a outras populações ciganas, de forma a complementar a informação obtida com os marcadores autossómicos e a produzir um conjunto de resultados mais amplo, a partir do qual se procurassem obter mais pistas sobre a tão pouco conhecida história dos ciganos.

4. Considerações finais e conclusões

A realização deste trabalho permitiu complementar alguns estudos já existentes sobre a história das comunidades ciganas, particularmente dos ciganos portugueses. Foi possível esclarecer alguns aspectos do seu complexo passado demográfico, obter achegas sobre os processos migratórios, as múltiplas fragmentações populacionais, assim como sobre as interações estabelecidas entre a população cigana portuguesa e a população circunvizinha.

A fim de traçar o perfil genético da população cigana de Portugal, em certa medida representativa dos ciganos da Península Ibérica, optou-se por analisar neste estudo marcadores moleculares localizados em autossomas e no cromossoma X. Os resultados obtidos nos ciganos portugueses foram comparados com os reportados para outras populações ciganas pertencentes maioritariamente ao Leste europeu, região que até agora têm sido o alvo preferencial de estudos de Genética Populacional envolvendo grupos Roma.

Este trabalho deu continuidade ao estudo iniciado por Gusmão *et al.* (2008), que, através do estudo de marcadores localizados no cromossoma Y, facultou uma primeira visão sobre os ciganos portugueses baseada no padrão de diversidade das linhagens masculinas. Procurou-se agora testar a coerência desses resultados através da perspectiva proporcionada por marcadores autossómicos e localizados no cromossoma Y.

Em conclusão, foram os seguintes os principais aspectos que esses marcadores evidenciaram:

- A baixa diversidade genética da população cigana portuguesa, comparativamente à população portuguesa não-cigana, bem visível ao nível dos dois tipos de microsatélites estudados. Tal como previamente reportado para outras populações ciganas europeias, esta observação reflecte a actuação de fortes efeitos de deriva genética, em parte consequente do baixo efectivo populacional dos grupos Roma. Por outro lado a homogeneidade intrapopulacional é reveladora das tradicionais práticas endogâmicas entre as populações ciganas.

- O grau elevado de diferenciação genética não só entre os ciganos portugueses e outras populações ciganas, mas em geral entre grupos Roma indiferentemente da distância geográfica que os separa. Esta heterogeneidade contrasta com a forte coesão entre populações europeias não-ciganas, amplamente reportada em diversos trabalhos.
- A tendência para as populações ciganas se agruparem entre si quando projectadas no contexto de outras populações europeias. Apesar de formarem um *cluster* pouco cerrado devido à grande diferenciação entre populações ciganas, o padrão de agrupamento revela um elevado grau de partilha do património genético entre os diferentes Roma da Europa, apontando claramente para a sua origem comum.
- A marcada diferenciação entre as populações ciganas e as respectivas populações hospedeiras, de ordem semelhante ao grau de diferenciação entre populações ciganas, mais uma vez revelador dos intensos efeitos de deriva que se fizeram sentir durante a história dos grupos Roma.
- A significativa contribuição não-indiana (europeia ou asiática) na formação das populações ciganas, por oposição ao componente ancestral indiano, aparentemente bastante diluído no *pool* genético dos actuais ciganos. A observação é consistente com os resultados obtidos por Gusmão *et al.* (2008) que também no cromossoma Y encontram evidência de considerável fluxo genético entre ciganos e não-ciganos. Ambos os resultados apontam para um grau de mistura entre populações ciganas e não-ciganas que, paradoxalmente, contraria a ideia comum que atribui aos grupos Roma uma natureza extremamente fechada. Por outro lado a relativa escassez de um componente ancestral com origem na Índia, pode indicar ter sido bastante reduzido o número de indivíduos indianos envolvidos na migração inicial.
- O padrão peculiar de desequilíbrio de ligação ao nível dos X-STRs, tradutor da complexa história demográfica dos ciganos portugueses, simultaneamente marcada por intensa deriva genética e mistura populacional. São factores, opostos em termos de efeitos, que parecem ter tido um papel fundamental na geração dos padrões de diversidade que caracterizam os ciganos da Europa.

5. Referências bibliográficas

Abecasis GR, Cookson WO. 2000. Gold – Graphical Overview of Linkage Disequilibrium. *Bioinformatics* 16:182-183.

Achim V. 2004. *The Roma in Romanian history* Budapest. CEU Press.

Amorim A, Alves C, Gusmão L, Pereira L. 2006. Extended Northern Portuguese database on 21 autosomal STRs used in genetic identification. *Progress in Forensic Genetics* 11. International Congress Series 1288:364-366. Elsevier Science. Amsterdam.

Anker R, Steinbrueck T, Donis-Keller H. 1992. Tetranucleotide repeat polymorphism at the human thyroid peroxidase (hTPO) *locus*. *Hum Mol Genet* 1:137.

Applied Biosystems, AmpF/STR Identifiler™ PCR Amplification *Kit* User's Manual. Foster City (CA): Applied Biosystems. 2005.

Çakir AH, Çelebioğlu A, Altunbaş S, Yardımcı E. 2003. Allele frequencies for 15 STR loci in Van-Ağrı districts of the Eastern Anatolia region of Turkey. *Forensic Sci Int* 135:60-63.

Balloux F, Goudet J. 2002. Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. *Mol Ecol* 11:771-783.

Bell GI, Jurka J. 1997. The length distribution of perfect dimer repetitive DNA is consistent with its evolution by an unbiased single-step mutation process. *J Mol Evol* 44:414-421.

Bernasovsky I, Suchy J, Bernasovska K, Vargova T. 1976. Blood groups of Roms (Gypsies) in Czechoslovakia. *Am J Phy Anthropol* 45:277-280.

Bertorelle G, Excoffier L. 1998. Inferring admixture proportions from molecular data. *Mol Biol Evol* 15:1298-311.

Bindu GH, Trivedi R, Kashyap VK. 2007. Allele frequency distribution based on 17 STR markers in three major Dravidian linguistic populations of Andhra Pradesh. India. *Forensic Sci Int* 170:76-85.

Budowle B, Brown BL. 2001. The use of DNA typing for forensic identification. *Forensica* 1:23-37.

Budowle B, Moretti TR, Niezgoda SJ, Brown BL. 1998. CODIS and PCR-based short tandem repeat *loci*: law enforcement tools. Proceedings of the Second European Symposium on Human Identification. Innsbruck. Austria. Madison.

Butler JM. 2001. *Forensic DNA Typing*. Academic Press. San Diego. CA.

Butler JM. 2006. Genetics and Genomics of Core Short Tandem Repeat *Loci* Used in Human Identity Testing. *J Forensic Sci* 51:2.

Calabrese PP, Durrett RT, Aquadro CF. 2001. Dynamics of microsatellite divergence under stepwise mutation and proportional slippage/point mutation models. *Genetics* 159:839-852.

Caglià A, Tofanelli S, Coia V, Boschi I, Pescarmona M, Spedini G, Pascali V, Paoli G, Destro-Bisol G. 2003. A study of Y-chromosome microsatellite variation in sub-Saharan Africa: a comparison between F_{ST} and R_{ST} genetic distances. *Hum Biol* 75:313-330.

Coelho FA. 1995. Os ciganos de Portugal com um estudo do calão. 1ª Edição na Dom Quixote (a partir da edição original de 1892). Lisboa.

Collins PJ, Hennessy LK, Leibel CS, Roby RK, Reeder DJ, Foxall PA. 2004. Developmental validation of a single-tube amplification of the 13 CODIS STR *loci*, D2S1338, D19S433, and amelogenin: the AmpF/STR Identifier PCR Amplification *Kit*. *J Forensic Sci* 49:1265–1277.

Cooperative Human Linkage Center (<http://www.chlc.org>) GATA7G07. 37564.

Cooperative Human Linkage Center (<http://www.chlc.org>) GATA3F01. 511

Cooperative Human Linkage Center (<http://www.chlc.org>) GATA7G10. 415.

Cooperative Human Linkage Center (<http://www.chlc.org>) GATA11C06. 715

Cooperative Human Linkage Center (<http://www.chlc.org>) GATA3F03. 512

Cooperative Human Linkage Center (<http://www.chlc.org>) GGAA3A09. 31762

Cooperative Human Linkage Center (<http://www.chlc.org>) GGAA2A03. 135

Costa, MA. 2003. Representações Sociais da Comunidade Cigana. Dissertação de Doutoramento em Antropologia, especialidade de Antropologia Social e Cultural. Faculdade de Ciências e Tecnologia. Departamentode Antropologia da Universidade de Coimbra.

Deligiannidis P, Triantaphyllidis C, Psaroulis D, Kouvatsi A. 2006. Forensic evaluation of 13 STR *loci* in the Roma population (Gypsies) of Greece. Forensic Sci Int 157:198-200.

Dupanloup I, Bertorelle G. 2001. Inferring Admixture Proportions from Molecular Data: Extension to Any Number of Parental Populations Mol Biol and Evol 18:672-675.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7:214.

Edelmann J, Hering S, Michael M, Lessig R, Deichsel D, Meier-Sundhausen G, Roewer L, Plate I, Szibor R. 2001. 16 X-chromosome STR *loci* frequency data from a German population. Forensic Sci Int 124:215-218.

Edelmann J, Deichsel D, Hering S, Plate I, Szibor R. 2002a. Sequence variation and allele nomenclature for the X-linked STRs DXS9895, DXS8378, DXS7132, DXS6800, DXS7133, GATA172D05, DXS7423 and DXS8377. Forensic Sci Int 129:99-103.

Edelmann J, Hering S, Kuhlisch E, Szibor R. 2002b. Validation of the STR DXS7424 and the linkage situation on the X-chromosome. Forensic Sci Int 125:217-222.

Edelmann J, Deichsel D, Plate I, Kayser M, Szibor R. 2003. Validation of the X-chromosomal STR DXS6809. *Int J Legal Med* 117:241-244.

Egyed B, Füredi S, Angyal M, Boutrand L, Vandenberghe A, Woller J, Pádár Z. 2000. Analysis of eight STR *loci* in two Hungarian populations. *Int J Legal Med* 113:272-275.

Excoffier L, Laval G, Schneider S. 2005. Arlequin ver 3.01: An integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47–50.

Fazito D. 2000. A identidade cigana e o efeito de “nomeação”: deslocamento das representações numa teia de discursos mitológico-científicos e práticas sociais. *Revista de Antropologia*. São Paulo. 49:2.

Felsenstein J. 1989. Phylip – Phylogeny Inference Package Version 3.2. *Cladistics* 5:164-166.

Fraser A. 1998. História do Povo Cigano. Tradução de Costa T. Editorial Teorema, LDA.

Gaggiotti OE, Lange O, Rassmann K, Gliddon C. 1999. A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Mol Ecol* 8:1513-1520.

Gaikwad S, Kashyap VK. 2003. Genetic diversity in four tribal groups of western India: a survey of polymorphism in 15 STR *loci* and their application in human identification. *Forensic Sci Int* 134:225-231.

Goldstein D, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW. 1995. Genetic absolute dating based in microsatellites and the origin of modern humans. *Proc Nat Acad Sci* 92:6723-6727.

Gomes I, Gomes I, Prinz M, Pereira R, Meyers C, Mikulasovich RS, Amorim A, Carracedo A, Gusmão L. 2007. Genetic analysis of three US population groups using an X-chromosomal STR decaplex. *Int J Legal Med* 121:198-203.

Gomes I, Prinz M, Pereira R, Bieschke E, Mayr WR, Amorim A, Carracedo A, Gusmão L. 2009a. X-chromosome STR sequence variation, repeat structure, and nomenclature in humans and chimpanzees. *Int J Legal Med* 123:143-149.

Gomes I, Pereira V, Gomes V, Prata MJ, Pinto N, Carracedo A, Amorim A, Gusmão L. 2009b. The Karimojong from Uganda: Genetic characterization using an X-STR decaplex system. *Forensic Sci Int Genet*. doi:10.1016/j.fsigen.2008.10.009 (*Epub ahead of print*)

Gomes V, Sánchez-Diz P, Alves C, Gomes I, Amorim A, Carracedo A, Gusmão L. 2009. Population data defined by 15 autosomal STR *loci* in Karamoja population (Uganda) using AmpF/STR Identifier *kit*. *Forensic Sci Int* 3:e55-e58.

Gresham D, Morar B, Underhill PA, Passarino G, Lin AA, Wise C, Angelicheva D, Calafell F, Oefner PJ, Shen P, Tournev I, de Pablo R, Kucinskas V, Perez-Lezaun A, Marushiakova E, Popov V, Kalaydjieva L. 2001. Origins and divergence of the Roma (gypsies). *Am J Hum Genet* 69:1314-1331.

Gusmão A, Gusmão L, Gomes V, Alves C, Calafell F, Amorim A, Prata MJ. 2008. A perspective on the history of the Iberian gypsies provided by phylogeographic analysis of Y-chromosome lineages. *Ann Hum Genet* 72:215-227.

Gusmão L, Sánchez-Diz P, Alves C, Gomes I, Zarrabeitia MT, Abovich M, Atmetlla I, Bobillo C, Bravo L, Builes J, Cainé L, Calvo R, Carvalho E, Carvalho M, Cicarelli R, Catelli L, Corach D, Espinoza M, García O, Malaghini M, Martins J, Pinheiro F, João Porto M, Raimondi E, Riancho JA, Rodríguez A, Rodríguez A, Rodríguez Cardozo B, Schneider V, Silva S, Tavares C, Toscanini U, Vullo C, Whittle M, Yurrebaso I, Carracedo A, Amorim A. 2009. A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: data on 15 Iberian and Latin American populations. *Int J Legal Med* 123:227-234.

Gyodi E, Tauszik T, Petranyi G, Kotvasz A, Palaffy G, Takacs NP, Hollan SR. 1981. The HLA antigen distribution in the Gypsy population in Hungary. *Tissue Antigens* 18:1-12.

Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R. 1994. Evaluation of 13 short tandem repeat *loci* for use in personal identification applications. *Am J Hum Genet* 55:175–189.

Hardy OJ, Charbonnel N, Fréville H, Heuertz M. 2003. Microsatellite Allele sizes: A simple Test to assess their significance on genetic differentiation. *Genetics* 163:1467-1482.

Havas D, Jeran N, Efremovska L, Dordević D, Rudan P. 2007. Population genetics of 15 AmpflSTR Identifiler *loci* in Macedonians and Macedonian Romani (Gypsy). *Forensic Sci Int* 173:220-224.

Hering S, Szibor R. 2000. Development of the X-Linked Tetrameric Microsatellite Marker DXS9898 for Forensic Purposes. *J Forensic Sci* 45:929-931.

Hering S, Kuhlisch E, Szibor R. 2001. Development of the X-linked tetrameric microsatellite marker HumDXS6789 for forensic purposes. *Forensic Sci Int* 119:42-6.

Hohoff C, Schürenkamp M, Borchers T, Eppink M, Brinkmann B. 2006. Meiosis study in a population sample from Afghanistan: allele frequencies and mutation rates of 16 STR *loci*. *Int J Legal Med* 120:300-302.

Jeffreys A, Wilson V, Thein SL. 1985. Hypervariable “minisatellite” regions in human DNA. *Nature* 314:67-73.

Jobling MA, Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nature Rev Genet* 4:598-612.

Jobling MA, Hurles ME, Tyler-Smith C. 2004. *Human Evolutionary Genetics - Origins, Peoples & Disease*. Garland Science.

Kalaydjieva L, Gresham D, Calafell F. 2001a. Genetic studies of the Roma (Gypsies): a review. *BMC Medical Genetics* 2:5.

Kalaydjieva L, Calafell F, Jobling MA, Angelicheva D, de Knijff P, Rosser ZH, Hurles ME, Underhill P, Tourné I, Marushiakova E, and Popov V. 2001b. Patterns of

inter- and intra-group genetic diversity in the Vlax Roma as revealed by Y chromosome and mitochondrial DNA lineages. *Eur J Hum Genet* 9:97-104.

Kalaydjieva L, Morar B. 2003. Roma (Gypsies): genetics studies. In: Cooper DN (ed) *Nature Encyclopedia of the human genome*. Vol 5. Nature Publishing Group. Londres. pp 160-165.

Kalaydjieva L, Morar B, Chaix R, Tang H. 2005. A newly discovered founder population: the Roma/Gypsies. *BioEssays* 27:1084-1094.

Katoh T, Mano S, Ikuta T, Munkhbat B, Tounai K, Ando H, Munkhtuvshin N, Imanishi T, Inoko H, Tamiya G. 2002. Genetic isolates in East Asia: a study of linkage disequilibrium in the X chromosome. *Am J Hum Genet* 71:395-400.

Kendrick D, Puxon G. 1998. Ciganos: Da Índia ao Mediterrâneo. Coleção Interface 3.

Kimpton CP, Walton A, Gill P. 1992. A further tetranucleotide repeat polymorphism in the vWF gene. *Hum Mol Genet* 1:287.

Kimura M, Crow JF. 1964. The number of alleles that can be maintained in a finite population. *Genetics* 49:725-738.

Kimura M, Ohta T. 1973. Eukaryotes-prokaryotes divergence estimated by 5S ribosomal RNA sequences. *Nat New Biol* 243:199-200.

Kruglyak S, Durrett RT, Schug MD, Aquadro CF. 1998. Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proc Natl Acad Sci USA* 95:10774-10778.

Lahn BT, Page DC. 1999. Four evolutionary strata on the human Y chromosome. *Science* 286:964-967.

Leite F. 2006. Análise da Estrutura Genética da População do Rio Grande do Sul através de Microssatélites Autossômicos e de Cromossomas Sexuais. Dissertação de Doutorado em Genética e Biologia Molecular. Instituto de Biociências. Universidade Federal do Rio Grande do Sul.

- Leite FP, Santos SE, Rodríguez EM, Callegari-Jacques SD, Demarchi DA, Tsuneto LT, Petzl-Erler ML, Salzano FM, Hutz MH. 2009. Linkage Disequilibrium Patterns and Genetic Structure of Amerindian and Non-Amerindian Brazilian Populations Revealed by Long-Range X-STR Markers. *Am J Phys Anthropol* 10 1002/ajpa 20998 (*Epub ahead of print*)
- Li H, Schmidt L, Wei MH, Hustad T, Lerman MI, Zbar B, Tory K. 1993. Three tetranucleotide polymorphisms for *loci*: D3S1352, D3S1358, D3S1359. *Hum Mol Genet* 2:1327.
- Li WH, Yi S, Makova K. 2002. Male-driven evolution. *Curr Opin Genet Dev* 12:650–656.
- Liégeois JP. 1989. Ciganos e Itinerantes: Dados socioculturais, Dados sociopolíticos, Santa Casa da Misericórdia de Lisboa. 1ªEdição.
- Machado FB, Medina-Acosta E. 2009. Genetic map of human X-linked microsatellites used in forensic practice. *Forensic Sci Int Genet* 3:202–204.
- Malayarchuk BA, Grzybowski T, Deremko MV, Czarny J, and Miscicka-Sliwka D. 2006. Mitochondrial DNA diversity in the Polish Roma. *Ann Hum Genet* 70:195-206.
- Marushiakova E, Popov V. 1997. Gypsies (Roma) in Bulgaria. Frankfurt am Main: Peter Lang.
- Mills KA, Even D, Murray JC. 1992. Tetranucleotide repeat polymorphism at the human alpha fibrinogen *locus* (FGA). *Hum Mol Genet* 1:779.
- Novokmet N, Pavčec Z. 2007. Genetic polymorphisms of 15 *AmpFI*STR *loci* in Romani population from Northwestern Croatia. *Forensic Sci Int* 168:43-46.
- Ohno S. 1967. Sex Chromosomes and Sex-linked Genes. Berlim. Springer.
- Pereira R, Gomes I, Amorim A, Gusmão L. 2007. Genetic diversity of 10 X chromosome STRs in northern Portugal. *Int J Legal Med* 121:192–197.

Perez RM. 1995. Os Ciganos da Portugal: O exótico cá dentro. *in* Coelho FA. Os ciganos de Portugal com um estudo do calão. 1ª Edição na Dom Quixote (a partir da edição original de 1892). Lisboa.

Pfaff CL, Parra EJ, Bonilla C, Hiester K, McKeigue PM, Kamboh MI, Hutchinson RG, Ferrell RE, Boerwinkle E, Shriver MD. 2001. *Am J Hum Genet* 68:198-207.

Polymeropoulos MH, Xiao H, Rath DS, Merrill CR. 1991. Tetranucleotide repeat polymorphism at the human tyrosine hydroxylase gene (TH). *Nucleic Acids Res* 19:3753.

Projić P, Skaro V, Samija I, Pojskić N, Durmić-Pasić A, Kovacević L, Bakal N, Primorac D, Marjanović D. 2007. Allele frequencies for 15 short tandem repeat *loci* representative sample of Croatian population. *Croat Med J* 48:473-477.

Rajkumar R, Kashyap VK. 2003. Evaluation of 15 biparental STR *loci* in human identification and genetic study of the Kannada-speaking groups of India. *Am J Forensic Med Pathol* 24:187-192.

Reyniers A. 1995. Gypsy Populations and Their Movements Within Central and Eastern Europe and Towards Some Oecd Countries. Organization for Economic CoOpertaion and Development, International Migration and Labour Market Policies Occasional Papers 1.

Ross MT *et al.* 2005. The DNA sequence of the human X chromosome. *Nature*. 434:325-337.

Sánchez-Diz P, Menounos PG, Carracedo A, Skitsa I. 2008. 16 STR data of a Greek population. *Forensic Sci Int Genet* 2:71-72.

Sarkar N, Kashyap VK. 2002. Genetic diversity at two pentanucleotide STR and thirteen tetranucleotide STR *loci* by multiplex PCR in four predominant population groups of central India. *Forensic Sci Int* 128:196-201.

Schaffner SF. 2004. The X Chromosome in population genetics. *Nat Rev Genet* 5:43-51.

- Sharma V, Litt M. 1992. Tetranucleotide repeat polymorphism at the D21S11 *locus*. Hum Mol Genet 1:67.
- Shin SH, Yu JS, Parkb SW, Minc GS, Chunga KW. 2005. Genetic analysis of 18 X-linked short tandem repeat markers in Korean population. Forensic Sci Int 147:35–41.
- Silva F, Pereira R, Gusmão L, Santos C, Amorim A, Prata MJ, Bettencourt C, Lourenço P, Lima M. 2009. Genetic Profiling of the Azores Islands (Portugal): data from 10 X-Chromosome STRs. Am J Hum Biol (*In press*)
- Sivakova D. 1983. Distribution of the three red-cell enzyme polymorphisms (ACP, PGM1 and AK) in gypsies from Slovakia (Czechoslovakia). Ann Hum Biol 10:449-452.
- Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457-462.
- Skaletsky H *et al.* 2003. The male-specific region of the human Y chromosome: a mosaic of discrete sequence classes. Nature 423:825-837.
- Soták M, Petrejčíková E, Bernasovská J, Bernasovský I, Sovicová A, Boronová I, Svicková P, Bôziková A, Gabriková D. 2008. Genetic variation analysis of 15 autosomal STR *loci* in Eastern Slovak Caucasian and Romany (Gypsy) population. Forensic Sci Int Genet 3:21-25.
- Straub RE, Speer MC, Luo Y, Rojas K, Overhauser J, Ott J, Gilliam TC. 1993. A microsatellite genetic linkage map of human chromosome 18. Genomics 15:48–56.
- Szibor R, Krawczak M, Hering S, Edelmann J, Kuhlisch E, Krause D. 2003a. Use of X-linked markers for forensic purposes. Int J Legal Med 117:67-74.
- Szibor R, Edelmann J, Hering S, Plate I, Wittig H, Roewer L, Wiegand P, Cali F, Romano V, Michael M. 2003b. Cell line DNA typing in forensic genetics—the necessity of reliable standards. Forensic Sci Int 138:37–43.
- Szibor R, Hering S, Edelmann J. 2006. A new Web site compiling forensic chromosome X research is now online. Int J Legal Med 120:252-254.

Venter JC *et al.* 2001. The Sequence of the Human Genome. *Science* 291: 1304-1351.

Waterston RH *et al.* 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520-562.

Weber JL, Wong C. 1993. Mutation of human short tandem repeats. *Hum Mol Genet* 2:1123-1128.

Wilson JF, Goldstein DB. 2000. Consistent long-range linkage disequilibrium generated by admixture in a Bantu-Semitic hybrid population. *Am J Hum Genet* 67:926-935.

Wren JD, Forgacs E, Fondon JW, Pertsemlidis A, Cheng SY, Gallardo T, Williams RS, Shohet RV, Garner HR. 2000. Repeat polymorphisms within gene regions: phenotypic and evolutionary implications. *Am J Hum Genet* 67:345-356.

Wright S. 1951. The genetic structure of populations. *Ann Eugenics* 15:323-354.

Zarrabeitia MT, Amigo T, Sañudo C, Pancorbo MM, Riancho JA. 2002. Sequence structure and population data of two X-linked markers: DXS7423 and DXS8377. *Int J Legal Med* 116:368-371.

Zavattari P, Deidda E, Whalen M, Lampis R, Mulargia A, Loddo M, Eaves I, Mastio G, Todd JA, Cucca F. 2000. Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: demography, chromosome recombination frequency and selection. *Hum Mol Genet* 20:2947-2957.

Anexos

Anexo 1.

A genetic historical sketch of European Gypsies. The perspective from autosomal markers

Alfredo Gusmão¹, **Cristina Valente**^{1,2}, Verónica Gomes^{1,3}, Cíntia Alves¹, António Amorim^{1,4}, Maria João Prata^{1,4}, Leonor Gusmão¹

¹*IPATIMUP, Instituto de Patologia e Imunologia Molecular da Universidade do Porto, 4200-465 Porto, Portugal*

²*Departamento de Biologia, Universidade de Aveiro, 3810-193 Aveiro, Portugal*

³*Institute of Legal Medicine, University of Santiago de Compostela E-15782, Santiago de Compostela, Spain.*

⁴*Faculdade de Ciências, Universidade do Porto, 4169-007 Porto, Portugal*

Abstract

In this work, 123 unrelated Portuguese Gypsies were analyzed for 15 highly polymorphic autosomal Short Tandem Repeats (STRs). Average gene diversity across the 15 markers was 76.7%, being lower than observed in the non-Gypsy Portuguese population. Subsets of STRs were used to perform comparisons with other Gypsies and corresponding host populations. Interestingly, diversity reduction in Gypsy groups comparatively to the non-Gypsy surrounding populations apparently varied according to an East-West gradient that parallels their dispersion in Europe as well as the decreasing complexity of their internal structure.

Analysis of genetic distances revealed that the average level of genetic differentiation between Gypsy groups was much larger than registered between the corresponding non-Gypsy populations. The high heterogeneity among Gypsies can be explained by strong genetic drift and limited inter-group gene flow. However, when genetic relationships were addressed through Principal Component Analysis, all Gypsy populations clustered together clearly differentiated from other populations, a pattern which testifies their common origin.

Concerning the putative ancestral genetic component, admixture analysis did not reveal a strong Indian ancestry in the current Gypsy gene pools contrasting to the high admixture estimates with either Europeans or Western Asians.

Anexo 2.

**The Portuguese Gypsy community: A genetic approach to the
demographic history provided by the analysis of autosomal markers**

Cristina Valente^{1,2}, Alfredo Gusmão¹, Verónica Gomes^{1,3}, Cíntia Alves¹, António Amorim^{1,4}, Maria João Prata^{1,4}, Leonor Gusmão¹

¹ *Institute of Molecular Pathology and Immunology of University of Porto (IPATIMUP), Portugal*

² *Department of Biology, University of Aveiro, Portugal*

³ *Institute of Legal Medicine, University of Santiago de Compostela E-15782, Santiago de Compostela, Spain.*

⁴ *Faculty of Sciences, University of Porto, Portugal*

Abstract

The main purpose of this study was to assess to the genetic history of the Portuguese Roma through the analysis of 15 autosomal microsatellites in a sample of 123 unrelated individuals. Average gene diversity for this battery of STRs was lower than the observed for the non-Gypsy Portuguese population showing the reduced genetic variability which characterises the Gypsy people. This feature results from random drift experienced during their diaspora as well from high levels of endogamy. Comparisons between Roma and corresponding host populations denoted that diversity reduction within Roma compared to the non-Roma followed a gradient of decrease according to a geographical direction quite similar to the route of migration and dispersion of Gypsy groups across Europe.

On one hand, when genetic distances were calculated high heterogeneity among the Roma groups was revealed, by greater genetic differentiation between Gypsy groups than between non-Gypsies populations, which can be explained again by the strong action of genetic drift and limited inter-group flow. Despite that, all Gypsy populations clustered together when genetic relationships were addressed by Principal Component Analysis clearly evidencing they share a common origin.

Admixture analysis was also conducted revealing a substantial European contribution to the current gene pool of all Roma groups, whereas the component ascribed to the ancestral origin in India comparatively appeared to be more diluted. Thus, it seems more likely that the incorporation of the Indian fraction autosomal genome stems primarily then European genetic contribution from contemporary times.

Anexo 3.

Association between STRs from the X chromosome in a sample of Portuguese Gypsies

Cristina Valente^{1,2}, Iva Gomes^{1,3}, Vânia Pereira^{1,4}, António Amorim^{1,4}, Leonor Gusmão¹, Maria João Prata^{1,4}

¹*IPATIMUP, Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Porto, Portugal*

²*Departamento de Biologia, Universidade de Aveiro, Portugal*

³*Instituto de Medicina Legal, Universidade de Santiago de Compostela E-15782, Santiago de Compostela, Spain.*

⁴*Faculdade de Ciências, Universidade do Porto, Portugal*

Abstract

In the present work 10 X-STRs (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902 and DXS6789) were analyzed in a sample of 123 unrelated males belonging to the Portuguese Gypsy community. A single Polymerase Chain Reaction (PCR) under multiplex amplification conditions was performed. Average diversity level among the Roma (0.721) was lower compared to the Portuguese non-Roma (0.745), and the genetic distance between Portuguese Gypsies and non-Gypsies was found to be statistically significant ($F_{ST}=0.019$; $p=0.000$).

Evaluation of the forensic efficiency of decaplex system in the Portuguese Gypsies indicated that the average power of discrimination was 0.999997 among males and 0.999999 among females. Combined mean exclusion chance in duos and trios was 0.999557 and 0.999988, respectively. In the Gypsy sample, analysis of pairwise linkage disequilibrium (LD) led to detect significant associations between DXS6809 and DXS6789 and between DXS6789 and DXS9898 (after Bonferroni's correction). The first two *loci* are in close linkage (0.35cM), but were not found to be associated in the non-Gypsy Portuguese sample. Although farther from each other DXS6789 and DXS9898 are at a distance of just 7.18 cM. Nevertheless, this is the first study where a significant association of these two pairs of markers is reported. In recently admixed populations, as the Gypsies are, high level of LD across *loci* might have important forensic implications.

This study reveals, for instance, that among the Portuguese Gypsies, associations between X-STRs should be considered in forensic applications instead of dealing independently with frequencies for each *locus*.